

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:25:21 ; Search time 28.82 Seconds
(without alignments)
3409.477 Million cell updates/sec

Title: US-09-662-812-2

Perfect score: 568
Sequence: 1 MGLFHLTFLGLCSLPISL.....TEGDAKNPVLGLIKIVE 568

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568	100.0	568	16	Q929G0 chlamydia p
2	12	2.1	566	16	084354 chlamydia t
3	10	1.8	566	16	09PK44 chlamydia m
4	9	1.6	240	2	Q9F3N4 streptomyces
5	9	1.6	408	2	Q9K529 pseudomonas
6	9	1.6	408	2	Q9AEY2 pseudomonas
7	9	1.6	425	16	Q9KJY9 rhizobium m
8	9	1.6	445	2	068141 rhodobacter
9	9	1.6	502	10	082752 arabidopsis
10	8	1.4	134	2	Q49007 mycoplasma
11	8	1.4	137	16	Q9PKF9 chlamydia m
12	8	1.4	173	17	097055 sulfolobus
13	8	1.4	197	16	097P95 streptococcus
14	8	1.4	206	2	Q54770 streptococcus
15	8	1.4	323	5	Q9B045 leishmania
16	8	1.4	337	11	Q91WE1 mus musculus

ALIGNMENTS

17	8	1.4	357	2	068120	068120 rhodobacter
18	8	1.4	411	2	051747	051747 pseudomonas
19	8	1.4	411	2	051851	051851 pseudomonas
20	8	1.4	411	2	09KAS1	09KAS1 pseudomonas
21	8	1.4	411	2	P95569	P95569 pseudomonas
22	8	1.4	412	1	Q9HR00	Q9HR00 thermococcus
23	8	1.4	414	2	056996	056996 zymomonas m
24	8	1.4	416	17	058762	058762 pyrococcus
25	8	1.4	434	16	092C76	092C76 listeria in
26	8	1.4	435	16	09WZJ3	09WZJ3 thermotoga
27	8	1.4	435	16	Q9KA24	Q9KA24 bacillus ha
28	8	1.4	435	16	Q9JUM0	Q9JUM0 staphylococ
29	8	1.4	447	16	Q9CG79	Q9CG79 lactococcus
30	8	1.4	448	16	Q99ZL9	Q99ZL9 streptococ
31	8	1.4	456	16	Q97R84	Q97R84 streptococ
32	8	1.4	477	10	Q9SEF4	Q9SEF4 arabidopsis
33	8	1.4	483	5	Q9XV67	Q9XV67 caenorhabd
34	8	1.4	509	2	Q93RV5	Q93RV5 streptomyce
35	8	1.4	558	10	Q93YN2	Q93YN2 arabidopsis
36	8	1.4	566	1	P77943	P77943 sulfolobus
37	8	1.4	570	10	Q948K1	Q948K1 citrus unsh
38	8	1.4	570	10	Q93X81	Q93X81 citrus para
39	8	1.4	587	16	Q9CGY7	Q9CGY7 lactococcus
40	8	1.4	596	5	Q950F9	Q950F9 helicoverpa
41	8	1.4	629	10	Q9FX95	Q9FX95 arabidopsis
42	8	1.4	629	10	Q39174	Q39174 arabidopsis
43	8	1.4	667	16	P71749	P71749 mycobacteri
44	8	1.4	998	5	O44800	O44800 caenorhabd
45	7	1.2	47	10	Q9M3R9	Q9M3R9 arabidopsis

RESULT	ID	PRELIMINARY	PRT	568 AA.
Q929G0	Q929G0			
AC	Q929G0			
DT	01-MAY-1999 (Trembl)	10, Created		
DT	01-MAY-1999 (Trembl)	10, Last sequence update		
DT	01-DEC-2001 (Trembl)	19, Last annotation update		
DE	PUTATIVE OMP PROTEIN.			
GN	CPN0021 OR CPJ0021 OR CP0755.			
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-CWL029;			
RX	MEDLINE=9920606; PubMed=10192388;			
RA	Kilman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,			
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis."			
RL	Nat. Genet. 21:385-389(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-AR39;			
RX	MEDLINE=20150255; PubMed=10684935;			
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,			
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,			
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,			
RA	Gavin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,			
RA	Eisen J., Fraser C.M.			
RT	"genome sequences of Chlamydia trachomatis Mopn and Chlamydia			
RT	pneumoniae AR39."			
RL	Nucleic Acids Res. 28:1397-1406(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-J138;			
RX	MEDLINE=20330349; PubMed=10871362;			
RA	Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,			
RA	Shiba T., Ishii K., Hattori M., Kunita S., Nakazawa T.,			
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138			

RT from Japan and CWL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL: AE001587; AAD18174.1; -
 DR EMBL: AE002234; AAF38558.1; -
 DR EMBL: AP002545; BAA98233.1; -
 DR TIGR: CP0755; -
 DR InterPro: IPR004155; HEAT_PBS.
 DR Pfam: PF03130; HEAT_PBS; 4.
 KW Complete proteome.
 SQ SEQUENCE 568 AA; 63553 MW; 3358FCB5BC482E80 CRC64;

Query Match 100.0%; Score 568; DB 16; Length 568;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLFHLTFLGLLCSPLSLVAKFPESVGHKILYISTOSTOQALATYLEALDAYGDHDF 60
 DB 1 MGLFHLTFLGLLCSPLSLVAKFPESVGHKILYISTOSTOQALATYLEALDAYGDHDF 60
 QY 61 VLKRGEDYLKOSTHSSDPQTRKSTIIGAGLAGSSEALDVSQAMETADPLQQLVLSAV 120
 DB 61 VLKRGEDYLKOSTHSSDPQTRKSTIIGAGLAGSSEALDVSQAMETADPLQQLVLSAV 120
 QY 121 SGHLGKTSDDLFLKALAPYPIRLAAYRLANKTKNTKVIDHLSFTIKLPEEIOCLSA 180
 DB 121 SGHLGKTSDDLFLKALAPYPIRLAAYRLANKTKNTKVIDHLSFTIKLPEEIOCLSA 180
 QY 181 IFLRLETESDAYIRDLAANKSAIRSATALQIGEYQOKRPLRLNLTSASPODOBAI 240
 DB 181 IFLRLETESDAYIRDLAANKSAIRSATALQIGEYQOKRPLRLNLTSASPODOBAI 240
 QY 241 LYALGKLDGOSYNNKKOLQKPDVDVTLAAQAALIALGKEEDALPVIKKQALEERPAL 300
 DB 241 LYALGKLDGOSYNNKKOLQKPDVDVTLAAQAALIALGKEEDALPVIKKQALEERPAL 300
 QY 301 YALRHLPSEIGIPALPFLTKNSEAKLVNALLLEIGCPTKLETERLVOPHNE 360
 DB 301 YALRHLPSEIGIPALPFLTKNSEAKLVNALLLEIGCPTKLETERLVOPHNE 360
 QY 361 TLALSFSKGRITLQNMKRVNIIVPODERERLSTTRGLEQIILFLRLKEAYLPCTY 420
 DB 361 TLALSFSKGRITLQNMKRVNIIVPODERERLSTTRGLEQIILFLRLKEAYLPCTY 420
 QY 421 KLASQKQTALTAISFISHTSHQDALDLFOAAKLPGEPIIRAYADIAYNLTKDPERK 480
 DB 421 KLASQKQTALTAISFISHTSHQDALDLFOAAKLPGEPIIRAYADIAYNLTKDPERK 480
 QY 481 RSLADYAKKLIQETLLFVDTEFNORPHSPMYLRYQVTPESRTKIMDLLETLATSKSED 540
 DB 481 RSLADYAKKLIQETLLFVDTEFNORPHSPMYLRYQVTPESRTKIMDLLETLATSKSED 540
 QY 541 IRLLIQMTGEDAKNFPYLAGLTIKIVE 568
 DB 541 IRLLIQMTGEDAKNFPYLAGLTIKIVE 568

RESULT 2
 ID 084354 PRELIMINARY; PRT; 566 AA.
 AC 084354;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE HYPOHETICAL 63.5 KDA PROTEIN.
 GN CF350.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 ON NCBI_TaxID-813;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/DW-3/CX;
 RX MEDLINE-99000809; PubMed-9784136;

RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatuzov R.L., Zhao Q., Koonin E.V.,
 RA Davys R.W.;
 "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis."
 RL Science 282:754-759(1998).
 DR EMBL: AE001308; AAC67945.1; -
 DR InterPro: IPR004155; HEAT_PBS.
 DR Pfam: PF03130; HEAT_PBS; 5.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 566 AA; 63507 MW; 92E5D4F190B23D56 CRC64;

Query Match 2.1%; Score 12; DB 16; Length 566;
 Best Local Similarity 100.0%; Pred. No. 0.0061;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 PYIRLEAAYRLA 152
 DB 139 PYIRLEAAYRLA 150

RESULT 3
 ID 09PK44 PRELIMINARY; PRT; 566 AA.
 AC 09PK44;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE HYPOHETICAL PROTEIN TC0629.
 GN TC0629.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 ON NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN / NIGG;
 RX MEDLINE-20150255; PubMed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Ueberback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwyn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE002331; AAF39458.1; -
 DR TIGR: TC0629; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 566 AA; 63143 MW; F9C7B8E7B4E1C6C2 CRC64;

Query Match 1.8%; Score 10; DB 16; Length 566;
 Best Local Similarity 100.0%; Pred. No. 0.61;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 IRLAAYRLA 152
 DB 141 IRLAAYRLA 150

RESULT 4
 ID 09F3N4 PRELIMINARY; PRT; 240 AA.
 AC 09F3N4;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE HYPOHETICAL FUSION PROTEIN.
 GN SC10F4.26C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomyces.

OX NCBI_TaxID=1902;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cereno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351: PubMed-8843436;
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinsahl H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL450350; CAC16986.1; -;
 DR InterPro: IPR004360; Gly_bleo_diox.
 DR Pfam: PF00303; Glyoxalase; 1.
 SQ SEQUENCE 240 AA; 2515 MW; EC8BA5C65D1D3DE6 CRC64;

Query Match 1.6%; Score 9; DB 2; Length 240;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LTSASPDQ 237
 DB 69 LTSASPDQ 77

RESULT 5
 ID Q9K529 PRELIMINARY; PRT; 408 AA.
 AC Q9K529;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE FERREDOXIN REDUCTASE.
 GN BPHA4.
 OS Pseudomonas sp. B4.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=59381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B4;
 RA Rodarie D., Willison J., Jouneau Y.;
 RT "Cloning, purification and molecular characterization of the biphenyl
 RT dioxygenase complex from Pseudomonas sp. strain B4.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -1-COFACITOR: FAD (BY SIMILARITY).
 DR EMBL: AJ251217; CAB93969.1; -;
 DR InterPro: IPR000759; Adnrx_reductase.
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR InterPro: IPR000205; NAD_binding.
 DR InterPro: IPR001100; Pyridine_redox_2.
 DR InterPro: IPR003042; Rng_moxxygenase.
 DR Pfam: PF00070; pyr_redox; 1.
 DR PRINTS: PR00419; ADXRDTASE.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00411; PNDRDTASEI.
 DR PRINTS: PR00469; PNDRDTASEII.
 DR PRINTS: PR00420; RINGMOXGNASE.
 KM FAD; Flavoprotein; Oxidoreductase.
 SQ SEQUENCE 408 AA; 42911 MW; 81F3BB10642D06B8 CRC64;

Query Match 1.6%; Score 9; DB 2; Length 408;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAGS 94
 DB 7 IIGAGLAGS 15

RESULT 6
 ID Q9AEY2 PRELIMINARY; PRT; 408 AA.
 AC Q9AEY2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE FERREDOXIN REDUCTASE.
 GN BPHG.
 OS Pseudomonas sp. Cam-1.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=85672;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CAM-1;
 RX MEDLINE-21268664; PubMed-11375179;
 RA Master E.R., Mohn W.W.;
 RT "Induction of bpha, Encoding Biphenyl Dioxygenase, in Two
 RT Polychlorinated Biphenyl-Degrading Bacteria, Psychrotolerant
 RT Pseudomonas Strain Cam-1 and Mesophilic Burkholderia Strain LB400.";
 RL Appl. Environ. Microbiol. 67:2669-2676(2001).
 CC -1-COFACITOR: FAD (BY SIMILARITY).
 DR EMBL: AY027651; AKI4785.1; -;
 DR InterPro: IPR000759; Adnrx_reductase.
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR InterPro: IPR000205; NAD_binding.
 DR InterPro: IPR001100; Pyridine_redox_2.
 DR InterPro: IPR003042; Rng_moxxygenase.
 DR Pfam: PF00070; pyr_redox; 1.
 DR PRINTS: PR00419; ADXRDTASE.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00411; PNDRDTASEI.
 DR PRINTS: PR00469; PNDRDTASEII.
 DR PRINTS: PR00420; RINGMOXGNASE.
 KM FAD; Flavoprotein; Oxidoreductase.
 SQ SEQUENCE 408 AA; 42982 MW; 975E2009AA387979 CRC64;

Query Match 1.6%; Score 9; DB 2; Length 408;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAGS 94
 DB 7 IIGAGLAGS 15

RESULT 7
 ID Q9KJY9 PRELIMINARY; PRT; 425 AA.
 AC Q9KJY9;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 2-OCTAPARENTYL-6-METHOXYPHENOL HYDROXYLASE (PUTATIVE OXIDOREDUCTASE
 DE PROTEIN).
 GN UBII OR SMC00316.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-1021;
 RX MEDLINE-20317087; PubMed-10858449;
 RA Schlenker C., de Rudder K.E.E., Kohrs V., Lopez-Lara I.M., Geiger O.,
 RT "Cloning and characterization of the gene for phosphatidylcholine
 RT synthase."
 RT J. Biol. Chem. 275:18919-18925(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021;
 RX MEDLINE-21368234; PubMed-11474104;
 RA Galibert F., Finan T.M., Long S.R., Puenhler A., Abola P., Ampe F.,
 RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
 RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
 RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
 RA Gloux S., Godtke T., Goffeau A., Golding B., Guzy J., Gurjal M.,
 RA Hernandez-Lucas I., Hong A., Hulzar L., Hyman R.W., Jones T., Kahn D.,
 RA Kahn M.L., Kallman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
 RA Masny D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
 RA Ramsperger U., Surzycki R., Thebaud P., Vandenbol M., Yeh K.-C., Batut J.,
 RA Vorholter F.J., Weidner S., Wells D.H., Wong K., Yen K.-C., Batut J.,
 RT "The composite genome of the legume symbiont Sinorhizobium meliloti."
 RL Science 293:668-672(2001).
 DR EMBL; AF155772; AAF27309.1; -
 DR EMBL; AL591788; CAC46252.1; -
 DR InterPro: IPR000733; Flavo_monooxygenase.
 DR Pfam: PF01360; Monooxygenase; 1.
 DR PRINTS; PR00420; RINGMONOXGNASE.
 KW Complete proteome.
 SQ SEQUENCE 425 AA; 45928 MW; E6EBCD75F0DE0FB7 CRC64;

Query Match 1.6%; Score 9; DB 16; Length 425;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 86 ITGAGLAGS 94
 |||||
 Db 30 ITGAGLAGS 38

RESULT 8
 ID 068141 PRELIMINARY; PRT; 445 AA.
 AC 068141;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE GID PROTEIN.
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
 NC Rhodospirillum.
 NCBI_TaxID=1061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SB1003;
 RX MEDLINE-97404404; PubMed-9256491;
 RA Vitek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fomstein M.,
 RT "Sequence of a 189-kb segment of the chromosome of Rhodospirillum
 RT capsulatus SB1003."
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
 DR EMBL; AF010496; AAC16231.1; -
 DR InterPro: IPR002218; GIDA.
 DR Pfam: PF003738; GIDA.1.
 SQ SEQUENCE 445 AA; 47890 MW; 1C699BDC1EDD24C4 CRC64;

Query Match 1.6%; Score 9; DB 2; Length 445;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 86 ITGAGLAGS 94

Db 6 ITGAGLAGS 14
 |||||
 RESULT 9
 ID 082752 PRELIMINARY; PRT; 502 AA.
 AC 082752;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 55.2 KDA PROTEIN.
 GN F7H9.220 OR AT4G23030.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Bancroft I.,
 RA Meves H.W., Mayer K.F.X., Schueller C., Beyan M.,
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Meves H.W.,
 RA Lemcke K., Mayer K.F.X.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL031018; CAI19819.1; -
 DR EMBL; AL161558; CAB79258.1; -
 DR InterPro: IPR002528; UPF0013.
 DR Pfam: PF01554; UPF0013; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 502 AA; 55232 MW; 3D5F8F6220B97BC4 CRC64;

Query Match 1.6%; Score 9; DB 10; Length 502;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 LLICSLPTS 19
 |||||
 Db 122 LLICSLPTS 130
 RESULT 10
 ID 049007 PRELIMINARY; PRT; 134 AA.
 AC 049007;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE SIMILAR TO NADH OXIDASE.
 OS Mycoplasma capricolum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Entomoplasmataceae.
 NCBI_TaxID=2095;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 27343(KID);
 RX MEDLINE-96059641; PubMed-7476192;
 RA Bork P., Ouzounis C., Casari G., Schneider R., Sander C., Dolan M.,
 RA Gilbert W., Gillevet P.M.,
 RT "Exploring the Mycoplasma capricolum genome: a minimal cell reveals
 RT its physiology."
 RL Mol. Microbiol. 16:955-967(1995).
 DR EMBL; Z33089; CAA83753.1; -
 SQ SEQUENCE 134 AA; 14933 MW; 759690580E8F8B0 CRC64;

Query Match 1.4%; Score 8; DB 2; Length 134;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAG 93
 |||||
 Db 7 IIGAGLAG 14

RESULT 11

Q9PKF9 PRELIMINARY: PRT; 137 AA.

AC Q9PKF9;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE CYCLIC NUCLEOTIDE-BINDING PROTEIN, PUTATIVE.
 GN TC0506.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN / NIGG;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA Linher K., Weidman J., Khouli H., Craven B., Bowman C., Dodson R.,
 RA Eissen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
 pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE002319; AAF39348.1; -;
 DR TIGR: TC0506; -;
 DR InterPro: IPR000595; CNMP_binding.
 DR Pfam: PF00027; CNMP_binding; 1.
 DR SMART: SMO0100; CNMP; 1.
 DR PROSITE: PS50042; CNMP_BINDING_3; 1.
 KW Complete proteome.
 SQ SEQUENCE 137 AA; 15410 MW; 607B533A70C9B7A7 CRC64;

Query Match 1.4%; Score 8; DB 16; Length 137;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 VALALLEL 338
 |||||
 Db 120 VALALLEL 127

RESULT 12

Q970S5 PRELIMINARY: PRT; 173 AA.

AC Q970S5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE PUTATIVE NADH-UBIQUINONE OXIDOREDUCTASE SUBUNIT K.
 GN S11527.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Ankal A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,

RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermoacidophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7."
 RL DNA Res. 8:123-140(2001).
 DR EMBL: AP000986; BAB6598.1; -;
 KW Ubiquinone: Hypothetical protein; Complete proteome.
 SQ SEQUENCE 173 AA; 19125 MW; E7D964FB948939A1 CRC64;

Query Match 1.4%; Score 8; DB 17; Length 173;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 LPSEIGIP 313
 |||||
 Db 129 LPSEIGIP 136

RESULT 13

Q97P95 PRELIMINARY: PRT; 197 AA.

AC Q97P95;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DE HYPOTHEICAL PROTEIN SPI1746.
 GN SPI1746.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacilli/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettein H., Nelson K.E., Paulsen I.T., Eissen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Raddue D.,
 RA Holtzapfe E., Khouli H., Wolf A.W., Uitterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae."
 RL Science 293:498-506(2001).
 DR EMBL: AE007467; AAK75822.1; -;
 DR TIGR: SPI1746; -;
 DR InterPro: IPR002819; HD.
 DR InterPro: IPR003607; HDC.
 DR Pfam: PF01966; HD; 1.
 DR SMART: SM00471; HDC; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 197 AA; 22548 MW; 7C6AD881F67805CB CRC64;

Query Match 1.4%; Score 8; DB 16; Length 197;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 LHDYAKKL 490
 |||||
 Db 57 LHDYAKKL 64

RESULT 14

Q54770 PRELIMINARY: PRT; 206 AA.

AC Q54770;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 22.4 KDA PROTEIN (FRAGMENT).
 OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=1140;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCC 7942;
 RA Phung L.T., Haselkorn R.;
 RT "Genes encoding the alpha subunit of carboxyltransferase of the
 acetyl-CoA carboxylase complex and GTP cyclohydrolase I from
 cyanobacterium Synechococcus sp. PCC 7942."
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U59236; AAB82044.1; -;
 DR InterPro: IPR003719; PhzC-PhzF.
 DR Pfam: PF02567; PhzC-PhzF; 1.
 KW Hypothetical protein.
 FT NON_TER 206
 SQ SEQUENCE 206 AA; 22386 MW; 9F2EE53736A97C00 CRC64;

Query Match 1.4%; Score 8; DB 2; Length 206;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BY 268 TLAAQAAL 275
 |||||
 DB 75 TLAAQAAL 82

RESULT 15
 ID Q9BJ45
 AC Q9BJ45; PRELIMINARY; PRT; 323 AA.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 36.6 KDA PROTEIN.
 OS Leishmania major.
 OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21256109; PubMed=11356515;
 RA Pedrosa A.L., Ruiz J.C., Tosi L.R.O., Cruz A.K.;
 RT "Characterization of three chromosomal ends of Leishmania major
 RT reveals transcriptional activity across arrays of reiterated and
 RT unique sequences."
 RL Mol. Biochem. Parasitol. 114:71-80(2001).
 DR EMBL: AF339905; AAK28283.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 323 AA; 36566 MW; 8C92C045309068DF CRC64;

Query Match 1.4%; Score 8; DB 5; Length 323;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ADPLQQL 115
 |||||
 DB 102 ADPLQQL 109

Search completed: August 20, 2002, 04:30:01
 Job time: 280 sec

RESULT	3				
AE001587/c					
LOCUS	AE001587	16448 bp	DNA	linear	ECT 01-DEC-2000
DEFINITION	Chlamydia pneumoniae section 3 of 103 of the complete genome.				
ACCESSION	AE001587	AE001363			
VERSION	AE001587.1	GI:4376271			

REFERENCE	AUTHORS	TITLE	JOURNAL	YEAR
1 (bases 1 to 16448)	Kalanian, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinberg, L., Grimwood, J., Davis, R. W. and Stephens, R. S.	Comparative genomes of <i>Chlamydia pneumoniae</i> and <i>C. trachomatis</i>	Nat. Genet. 21 (4), 385-389 (1999)	1999

REFERENCE	4 (bases 1 to 16448)
AUTHORS	Kalman, S., Mitchell, W.,
TITLE	Grimwood, J., Davis, R.W. and Stephens, R.S.
JOURNAL	Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 225 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES	Location/Qualifiers

source . 1. .16448

```
misc_feature    172.    .2957
```

CONCERNING LIAISON SILLI
173 705

gene 615. .2957

gene 3196. .4971

```
misc_feature 3196.  .5953
```

gene	4874.5953
------	-----------

gene . 6302. .7786

```

misc_feature    6302.  .9139

```

gene 7787. .9139

gene complement(9380. .11467)

```

CDS      complement(9380. .11467) .

```

```
/transl_table=11
```

```
/db_xref="GI:4376272"
```

```

/product_id="A018173.1" /leader peptide: (outer membrane)
/protein_id="A018173.1"
/db_xref="GI:4376272"
/translation="MNRKSLFLASFLMGSSADALTHQEAVKKNKSYLSHRKSVGIVIT
TIDEGYLIINNNLRIDIANNVYVTEVGGSLIKVANGNNVNVRAATLEQVLEVEDT
DSCLLTNGRPAAMPGLGQSMYTLNPTETVIRKQVISTSEGGKRLCSGGYLESTV
SLSLGKTIETLRVCRDILPLPRPSINPMPIKPPVNFNGVGGFLAGTLYGASVPSIR
SLHSSFTFLDFSPFKHGVGQFNHSCQKQPPNPVNMKSYVNHRLAIDMAAHPHRSR

```

HGDCEPFAKNEGSEGLHSDMEIYADIFPNMFKMLGTPRDCNDTNDYEGVYLT
 SVSYVNFQANONEPLATLROYISITVMYGLAENIYECGLTACDFHDIWVGFSSP
 LLRAPIRLKHYVTLPGITSLSTGSLIYSDVPEISRSLSOLAKTOLDYREPLAKR
 SYTORHRIETPVYLTETPEPLAKNDNDHYIETSDAPSHISINAKRAGDITSTHNR
 PPRHAKLMTHTLHLSNESEPTPTKACELSPFGKATVYSIDAEIMIKKCMQHNRI
 RMWIEPCDMWNTLSESHKRSISLICKENEDTLVDSRIDLDDLSPLSDNRNILLK
 LEVPEHPCDMWNTSLRAGKROTPYLTOMILTKLIFEHMOYGVYEREADNSRFP
 FFLKLDKPKKPP
 complement(11484, .13190)

complement(11484. .13190)

```
/transl_table=11
```

```
/db_xref="GI:4376273"
```

QAMETADPLQQLLVLSAVSGHLGKTSDDLFLKALASPYPVIRLEAYRLANLKNTKV
 DHHSFETIHK DEETOCI SAATEP BEETESDAYTBN IAAKVSATBSAENIOTCEVOC

ALGKEEDALV I KQAL EERPRALVALRHLPS EIGIPALPIFLKTRNSEAKLNVALA
 I I E I G C D T B K I A E Y T E B I V O P H Y N E T I A I S E S K G B T I O N K R V N I T Y B D B O E P E B I

MPYLRYOVTPESEBTKLM.DT.ETTATSKSSEDI R.I.IOTMTEGDAKNBPYIAGILIKT
EQAALPGEELIKAYADLAIYNLIKDPEKKRSLHYAKKLIQEILLEVVDIENORPHPS

```
complement(131/2. .13/04)
/gene="maf"
```

```
comp+emmc(101/2:10/02)
/gene="maf"
```

```
/product="Maf-type protein"
```

```

/translation="MSLPLVLGSSSPRRKFILEKFRVPFTVIPSNFDESKVSYSGDPII

```

ILKKVHGCYNVQGLPIQTLKYLLE

/note="CPn0023"

```
/codon_start=1
```

```

/protein_id="AAD18176.1"

```

LKIMGMIEPTRGSI^{SL}PKKVGI^{LR}

HIADIDYDTI IYPGNYDDMVEKKT

LADCGQETLEWLNRKTGINDEI

complement(15438, .16376)

complement(15438. .16376)

```
/transl_table=1
```

/cd_xiei-SI:43/02/0
/translation="MASIYSFLDYLMVKSASPHTLRNYCLDNGIKITLFRGNJA

THIS PAGE BLANK (USPTO)

PSSPQLATEKRVSLPFLSFTKEHVMTIAKLIENGKAKRTIKRLSLSIKFAHYC
 VIOKILLENPAETIHGPRLPELSPMTYAOVETAMTPDISKJGJLDRCLMELFYS
 SGLRISEIVANODFDLSTFLIRIRGKGERIIPYTSNAIOIOLYLNRPDKRLE
 KDPAIFLNRGRIRISIDRSFOEYLRBSGLSHLPHIRHTIAFHMESEMDLK
 TIOALHSHSELTYYTQVSVKAKKOTRDBAHPHA*
 BASE COUNT 4710 a 3330 c 3665 g 4743 t
 ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 16448;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtcgctgatttggaaaagtcgagtggtgttataatgtcaagctccatcc 60
 Db 13290 GTGGCTGATTGAAAAAGGTCATGATGTGTTATATATTGAAGTCCCTATCC 13231
 Qy 61 aaacattgaactctgttagagaggttgacatgcatctatggtgactatccatcaac 120
 Db 13230 AAACATTGAATCTGTAGAGAGAGTGAACATGATCTATGGACTATTCATCTAAC 13171
 Qy 121 tctcttggaactttattgtgtagctctccatctctctgtgtgtaattccctgagtc 180
 Db 13170 TCTCTTGGACTTTATTGTGTAGTCTCCATTTCTGTGTGCTAAATTCCTGAGTC 13111
 Qy 181 tgaagtcataagatcccttatataagtcacatctacacagcagccttagaacaata 240
 Db 13110 TGAAGTCATTAAGATCCTTATATATAGTACGCAATCTACACAGCAGCTTAGCAACATA 13051
 Qy 241 tctggaagctctagatgctacggtgatacgaactctcgttcttaagaaaaatcgaga 300
 Db 13050 TCTGGAAGCTTATAGTCTAGAGTATCATGACTCTTGTTTAAAGAAAAATCGAGA 12991
 Qy 301 agactatctcaagcaaaagatccactccctcagatccgcaacttagaaaaagacatcat 360
 Db 12990 AGACTATCTCAAGCAAAAGATCCACTCCCTCAGATCCGCAAACTAGAAAAAGACCATCAT 12931
 Qy 361 tgaagcagcctgagcgagatcttcaagaagccttgagcgtcctcccaagcatctgaac 420
 Db 12930 TGAAGCAGCCTGAGCGGATCTTCAGAAAGCCTTGAGCGTCTCCCAAGCATRAGAAAC 12871
 Qy 421 tgcagacccctgacagcactagtttatacgagcgtcagacatcttgggaaaac 480
 Db 12870 TGCAGACCCCTGACAGCCTAGTGTATTCGCGCAGCTCAGACATCTTGGGAAAAC 12811
 Qy 481 tctgagcgaactctgtttaaagctttagatctcccttccctgtatccgttagaac 540
 Db 12810 TTCTGACGACTACTGTAAAGCTTTAGCATCTCCCTATCCGTATCCGCTTAGAAGC 12751
 Qy 541 cgcctatagactgtcgaatttgaagaacataagtcattgcatctatcttcat 600
 Db 12750 CGCCTATAGACTGTCAATTTGAAGAACTAAAGTCATTCATCATCTTTCTTCAAT 12691
 Qy 601 tcaatagctcccgaaagaatccaatgctatctcggaataatctcagcttggagac 660
 Db 12690 TCATTAAGCTTCCCGAAGAAATCCAAATGCTATCTGCGCAATATCTTAAGCTTGAAGAC 12631
 Qy 661 tgaagaatctgagcttatattcgagatctcctagctgcagaagaagcggaattggag 720
 Db 12630 TGAAGAATCTGAGCTTATATTCGGGATCTTAACTGCGCAAAAAAGCGGATTCGGAG 12571
 Qy 721 tgcacagacttgcagatcggaataacacacaaaaagcttctccgcagacttagaa 780
 Db 12570 TGCACAGCTTTCAGATCGGAGAAATACACAAAAAGCTTCTTCCGACATTTAGGA 12511
 Qy 781 tttgctaaagatgctctcctcctaagaatacaagactatctttagctttaggaagct 840
 Db 12510 TTTGCTAAAGATGCGCTCTCTCAAGATCAAGAAAGCTATCTTATGCTTAGGAAGCT 12451
 Qy 841 taagagatgctcagagctactacataataaaaaagcaattgcagaagccttgatgt 900
 Db 12450 TAAAGATGCTCAGAGCTACTACATATATAAAAAAGCAATTCAGAAAGCCTGATGTGATGT 12391

Qy 901 caattagcagcgtcagactttaatgtcttgggaaaagagagcgtcttccgt 960
 Db 12390 CACTTAGCAGAGCTCAAGCTTAAATGCTTTGGGAAAAAGAGAGAGCGCTTCCCGT 12331
 Qy 961 gataaaaaagcagcacttggagcggtccctcgagccctgtatgctttaggaatcacc 1020
 Db 12330 GATAAAAAAGCAGCACTTGAGAGCGGCTCGAGCCCTGTATCTTACGCACTTACACC 12271
 Qy 1021 ctctgagaatggatctcgacttgccttcgcaatctcctaaagaacagcagaagc 1080
 Db 12270 CTCTGAGTATGGGATTCGATTCGCCCTCGCATATCTTAATAAACTAAGACAGGAGAC 12211
 Qy 1081 caagttgaatgacttagctctctcttagagttgaggtgtgacacccctaaactactga 1140
 Db 12210 CAAGTTGAATGATGAGCTTAGCTCTTATAGATGAGGTGTAGACACCCCTAAACTACTGGA 12151
 Qy 1141 atacaatccgaagaagcttgcacaacacatataatgagacttagccttgagttctc 1200
 Db 12150 ATACATTAACGGAAGCTGTGTCACACCACTTATATAGACTGTAGGCTTGAAGTTCTC 12091
 Qy 1201 taagggcgctactttagcaaaattggaacgggtgagacatgaatccctcaagatccca 1260
 Db 12090 TAAAGGCGCTACTTATCAAAATTTGAGAGCGGCTGAACATCATAGTCCCTCAAGATCCCA 12031
 Qy 1261 ggaagggaaaaggttgcctccacacacccgaggtcttgaagagcagatcccttagct 1320
 Db 12030 GGAGAGGGAAGAGTGTGCTCTCCACAAACCGAGGCTTGAAGAGCAGATCTTACGTTCT 11971
 Qy 1321 ctccgcctacctaagaagcttaccctccctgtatataagcttggcgagtcagaa 1380
 Db 11970 CTTCGCGCTACTTAAAGAGCTTACCTCCCTGTTTATTAAGCTTTGGGAGTCAANA 11911
 Qy 1381 aactcagcttgcacactagcagatctcttttaagtcacacactacacaaagaaagcct 1440
 Db 11910 AACTCAGCTTGCACACTAGCTGCGATTTCTTTTAAAGTCAACACTTACATCAGAGAGCCTT 11851
 Qy 1441 agatcacttctccaagcttgcgaagcttccctgtagaagacatcatccgcgcctatgcaga 1500
 Db 11850 AGATCACTTCTCCAAGCTGCGAGCTCTCTGAGAACCTATCATCCGCGCTATGACAGA 11791
 Qy 1501 tcttgcctatctaatctcaacaaagatccctgaaaaaaagcttctccatgatatgc 1560
 Db 11790 TCTTGCCTATTTTAAATCTCACCAAAAGATCTGAAAAAAACCTTCTTCCATGATTATGC 11731
 Qy 1561 aaaaaagcctaattcaggaagacctttagtttggacaaggaagacaaagaagcccatcc 1620
 Db 11730 AAAAAAGCTAATTCCAGAAACCTTGTATTGTGACACGAGAAACCAAAAGCCCATCC 11671
 Qy 1621 cagcatgcccctactctacgttatacaggtcaccccaagaagcgttacgaagctcatgtgga 1680
 Db 11670 CAGCATGCCCTATCTACGTTATCAGGTACCCACAGAAAGCCGTAGAGCTCATGTGGA 11611
 Qy 1681 tattctagaagacactagcactcgaggtcttcggaagataccgtttatgatacaact 1740
 Db 11610 TATTCTAGAGACACTAGCACCTCGAAGCTTCCGAAGATATCCGTTATTTATATCAACT 11551
 Qy 1741 gatgacgaagagatgacaaaaaatttcccaagcttcccttggaagcttactcataaaatgt 1800
 Db 11550 GATGACGAGAGAGATGCAAAAAATTTCCCAAGCTCTTGCAGGCTTACTATAAAATTTGT 11491
 Qy 1801 ggaagtaaccccaactagctcttatagaacgltgcttcttattcttagcttccctgttc 1860
 Db 11490 GGAGTAACCCCAACTAGCTCTTATGAACGTTGCTTATTATTCAGCTTCTTGTTC 11431
 Qy 1861 ttatggttccctcagctgagttgtttagcttcaatcaagaagctgtgaa 1907
 Db 11430 TTATGTTCTTCAGCTGATGCTTGACATCAAGAGGCTGTGAAA 11384

RESULT 4
 AX349527 1707 bp DNA linear PAT 06-FEB-2002
 LOCUS AX349527
 DEFINITION Sequence 50 from Patent WO0202606.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:10:16 ; Search time 66.94 Seconds
(without alignments)
1467.898 Million cell updates/sec

Title: US-09-662-812-2
Perfect score: 2833
Sequence: 1 MGLEHLTLFGLLCSLPISL.....TEGDKNPVLGLIKIVE 568

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SEPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2833	100.0	568	16	Q929G0 chlamydia p
2	1456	51.4	566	16	Q9PK44 chlamydia m
3	1421	50.2	566	16	Q84354 chlamydia t
4	136.5	4.8	514	5	Q9UB78 drosophila
5	134.5	4.7	514	5	Q76522 drosophila
6	134	4.7	860	3	P87295 schizosacch
7	133	4.7	321	16	Q91112 pseudomonas
8	133	4.7	1138	16	Q84133 chlamydia t
9	132.5	4.7	514	5	Q9V455 drosophila
10	132.5	4.7	948	4	Q9UK7 homo sapien
11	132.5	4.7	1200	17	Q9HKM6 schizosacch
12	132.5	4.7	2554	2	Q30981 bacillus su
13	131	4.6	398	2	Q44295 anabaena sp
14	130.5	4.6	979	11	Q9EP71 mus musculu
15	129.5	4.6	1316	10	Q9M6P9 arabidopsis
16	129.5	4.6	1973	3	Q9Y7K2 schizosacch

17	129.5	4.6	7201	5	Q9VMT8 drosophila
18	127.5	4.5	1978	4	Q15154 homo sapien
19	126.5	4.5	1292	10	Q9SL02 arabidopsis
20	124.5	4.4	526	16	Q9PK10 chlamydia m
21	124	4.4	712	11	Q9CRX4 mus musculu
22	124	4.4	1211	11	Q9R0L3 rattus norv
23	124	4.4	1235	11	Q9R0L5 rattus norv
24	124	4.4	1273	11	Q9R0L4 rattus norv
25	124	4.4	1275	3	Q9PE64 neurospora
26	123.5	4.4	890	10	Q9M649 arabidopsis
27	123.5	4.4	894	10	Q9S0S3 arabidopsis
28	123.5	4.4	894	10	Q9M650 arabidopsis
29	123.5	4.4	1220	3	Q9P3A8 schizosacch
30	123.5	4.4	1591	4	Q92601 homo sapien
31	123.5	4.4	1723	2	Q9JMX8 heliobacte
32	123.5	4.4	1979	4	Q9S549 homo sapien
33	122	4.3	719	10	Q81742 arabidopsis
34	122	4.3	1318	4	Q43254 homo sapien
35	121.5	4.3	1841	4	Q15031 homo sapien
36	121	4.3	596	16	Q9RVT5 deinococcus
37	121	4.3	846	10	Q9EP99 oryza sativ
38	120.5	4.3	430	17	Q9UT9 pyrococcus
39	120.5	4.3	937	11	Q9DBG3 mus musculu
40	120.5	4.3	951	4	Q96019 homo sapien
41	120	4.2	922	16	Q9HMA7 pseudomonas
42	120	4.2	1388	6	Q9GL21 canis fami
43	119.5	4.2	669	4	Q96M49 homo sapien
44	119.5	4.2	1466	10	Q9ZCO3 arabidopsis
45	119.5	4.2	1588	11	Q9ESK9 mus musculu

ALIGNMENTS

RESULT	ID	Q929G0	PRELIMINARY;	PRT;	568 AA.
AC	Q929G0				
DT	01-MAY-1999	(TEMBLrel. 10, Created)			
DT	01-MAY-1999	(TEMBLrel. 10, Last sequence update)			
DT	01-DEC-2001	(TEMBLrel. 19, Last annotation update)			
DE	POTATIVE OMP PROTEIN.				
GN	CNP0021 OR CPJ0021 OR CP0755.				
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=83558;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CWL029;				
RC	MEDLINE=9206606; PubMed=10192388;				
RA	Kelman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,				
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,				
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";				
RL	Nat. Genet. 21:385-389(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=AR39;				
RC	MEDLINE=20150255; PubMed=10684935;				
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,				
RA	White O., Hickey E.K., Peterson J., Ueberlack T., Berry K., Bass S.,				
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,				
RA	Gavin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,				
RA	Eisen J., Fraser C.M.;				
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia				
RT	pneumoniae AR39.";				
RL	Nucleic Acids Res. 28:1397-1406(2000).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=J138;				
RC	MEDLINE=2030349; PubMed=10871362;				
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,				
RA	Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;				
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138				

RC STRAIN-D/UM-3/CX;
 RA MEDLINE-9900809; PubMed-9784136;
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.,
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis";
 RT Science 282:754-759(1998).
 DR EMBL: AE001308; AAC67945.1;
 DR InterPro: IPR004155; HEAT_PBS.
 DR Pfam: PF03130; HEAT_PBS; 5.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 566 AA; 63507 MW; 92E5D4F190B23D56 CRC64;

Query Match 50.2%; Score 1421; DB 16; Length 566;
 Best Local Similarity 53.0%; Pred. No. 6.1e-86;
 Matches 301; Conservative 88; Mismatches 177; Indels 2; Gaps 1;

QY 1 MGFLHTLFGILCSPLSLAKFPESVGHKILYISTOSTQALATVYLAADAYGDHDF 60
 DB 1 MGSLRLAFISFL--SFTLSACDFPSSVSQRILFSCSKSVPALEATLEASATYQCHDF 58
 QY 61 VLRKIEDYIKOSIHSSDPQTRKSTIIAGLAGSSSEALDVLQAMETADPLQQLVLSAV 120
 DB 59 VLRVIESYIQQSFLESDYIRKSAIIAGLSSSEALDVLQAMETADPLQQLVLSAV 118
 QY 121 SGHLGTSDDLFLKALASPVYIRLEAAYRLAKNTKYVDHLSFHFKRPEEIOCSAA 180
 DB 119 TQSLSTSDLLFKGLTASHPIRLAAYRLAKMKSVDLYSFYKPEEIOCSAA 178
 QY 181 IFRLTEESDAYIRDLAKKSAIRSATLQGEYQKRFPLRLRLTSLASPODEAI 240
 DB 179 IFQLETEEDAIHHLSSPNLTRYAVYLIGEKQKFLPLSLSLASPLDDEGA 238
 QY 241 LVALGKLKDGQSYNNIKQLOKPDVYTLAAQALIALGKEEDALPVYKQALEERPRL 300
 DB 239 LVALGKLEDSGYRPIKALSSRSNPVYTLAAQTLFLKEEELPLITNLCOQKLRLAL 298
 QY 301 YARHLPSELGIRIAPLPIKTKNSEKLVNALLLEGCDTRKLEYTERLYQPYNE 360
 DB 299 YARFELSQEGEELLPIFNATQETIRNTALALVHQGTDPQVLYLLEIESKVLV 358
 QY 361 TLALESKGTLLQMKRNVIIVPODERERLSTTGLEOILTFRLPKREAYLPCY 420
 DB 359 ILEPHTSTGAIOPKKECTTFPLMSQEDKMTAMRYVAADTILSLKLPPNDAYLYLE 418
 QY 421 KLLASQKQTLATTAISFLSHTSHQEAALLFOAKLPEDEPIRAYADLAIYNTKDEPK 480
 DB 419 RLASQKTLIAKALIAFLSVTAHQALSLYSKALFPQDPIRAYANLAIYMTKDEPK 478
 QY 481 RSHADAKKILQETLFLVTEENORPHSPMYLKYQVPESEKLMIDILETATSKSED 540
 DB 479 AVLYRAAEQIEDTILFETAENLEPSSSSYLRYOVSPETRTQMLALILELTVSSKTED 538
 QY 541 IRLLIQTEGDAKNFVLAGLITIKYE 568
 DB 539 IRYFLSMKTKHYKNPIISGLMRYVE 566

RESULT 4
 Q9UB78 AC PRELIMINARY; PRT; 514 AA.
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE IMPORTIN ALPHA-3 (IMPORTIN ALPHA 3).
 GN KARYOPHERIN-ALPHA-3 OR KARYOPHERIN-ALPHA3 OR CG9423.
 OS Karyopherin melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mathe E., Bates H., Huikeshoven H., Glover D., Cotterill S.,
 RT "Importin- α 3 is required at multiple stages of Drosophila development
 RT and has a stage specific role in the completion of oogenesis";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Mathe E., Cotterill S.,
 RT "Importin- α 3 cdna (gm06753), Drosophila melanogaster";
 RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Mathe E., Cotterill S.,
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ237997; CAB40789.1;
 DR EMBL: AF230873; AAF37856.1;
 DR EMBL: AF230872; AAF37855.1;
 DR EMBL: AF230871; AAK14941.1;
 DR HSSP: 002821; 1BK5.
 DR Flybase: FBgn0027338; karyopherin- α 3.
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 8.
 DR Pfam: PF01749; IB; 1.
 DR SMART: SM00185; ARM; 6.
 DR PROSITE: PSS0176; ARM_REPEAT; 2.
 SQ SEQUENCE 514 AA; 57036 MW; 5F007D9C73B85B8F CRC64;

Query Match 4.8%; Score 136.5; DB 5; Length 514;
 Best Local Similarity 21.1%; Pred. No. 0.52; Length 179; Indels 155; Gaps 23;
 Matches 108; Conservative 70; Mismatches 179; Indels 155; Gaps 23;

QY 95 SEALDVLQAMETAD---PLQQLVLSAVSGHLGTSDDLFLKALASPVYIRLEAAYRL 151
 DB 62 SSSIDLKAKKAADATKPEQQLAA-----YQAAKRL 93
 QY 152 ANL-KNTKYVDHLSFHFKRPEEIOCSAIFLRLTEESDAYIRDLAKKSAIRSATA 210
 DB 94 LSLDKNPINDLIQSDI--PLIIVELCKOHNTMLQFEAAWAL-----TNASGTS 142
 QY 211 LQIGEYQKRFPLRLNLTLASPODEALILYAGK-----LKDGQSYNNIKQLO-- 261
 DB 143 AQINEVVAAGAVPLFDLNSPAPNCEQAWALGNIIIGDGPLRDYVYKGVQPLLSF 202
 QY 262 -KPDVDTLA--AAQALIALGKEEDALPVYKQALEERPRLALRH----- 305
 DB 203 IKRPDIPITFLRNVTWVYVNLCKRNKDPAP--PTATIHETLRALNLTHTDNLIVDYWA 260
 QY 306 -----LPSEIG-IPLALPIFLKTKNSEAKLVN---ALALLEGCDT----- 342
 DB 261 ISYLTLDGNDQIQWIVIESGVVFKLIPLL--GNSEVYQVTAALRAVGNITVGSDEQOVV 317
 QY 343 -----PKLEYITERLYQVPHYNETLASEKGTLLQMKRNVIIVP----- 383
 DB 318 LNYDALSYFPGLLSHPEKTIKE-----AWFLSNITAGQSOVQAVINVGLLPKIEN 371
 QY 384 -----ODPDERERLRS--TTRGLEOILTFRLPKREAYLPCYIKKLSQKQTLATTAI 435
 DB 372 LSKGEFQTKKAAWALINLTISGREGVFTLL--KRGVIPPCCDLLSCQDTQVINYVL 427
 QY 436 SFLSH-----TSHQEA-LDLFOAKLPEDEPIRAYADLAIYNTKDEPKRSLLHYAKK 489
 DB 428 DGLNLMLKVDASHVEAVANCEECEGLAKTERLOSHEVETLKL-----AYE 474
 QY 490 LIOETLFLVDT-ENQRPSPMYLRYQVTPES 520
 DB 475 IILDQ--YFTDGEQOTNMAPSSDGAQVNFDPHA 504

RESULT 5

ID	076522	PRELIMINARY;	PRN:	514 AA.
AC	076522;			
DT	01-NOV-1998	(TREMBLrel, 08, Created)		
DT	01-NOV-1998	(TREMBLrel, 08, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel, 17, Last annotation update)		
DE	KARYOPHERIN ALPHA 3 (FRAGMENT).			
GN	KARYOPHERIN-ALPHA-3 OR CG9423.			
OS	Drosophila melanogaster (fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidae; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Mason D.A., Goldfarb D.S.;			
RT	"Drosophila melanogaster karyopherin alpha 3.";			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF074958; AAC26056.1; -.			
DR	HSSP; Q02821; 1BK5.			
DR	FLYBase; FBgn0027338; karyopherin-alpha-3.			
DR	InterPro; IPR000225; Armadillo.			
DR	InterPro; IPR002652; IBB.			
DR	Pfam; PF00514; Armadillo_seg; 8.			
DR	Pfam; PF01749; IBB; 1.			
DR	SMART; SM00185; ARM; 6.			
DR	PROSITE; PSS0176; ARM_REPEAT; 2.			
FT	NON TER			
SO	SEQUENCE	514 AA; 56975 MW; 78500AC4EA906D3A CRC64;		

Query Match	4.7%	Score 134.5	DB 5	Length 514
Best Local Similarity	20.9%	Pred. No. 0.71		
Matches 107; Conservative	69;	Mismatches 182;	Indels 153;	Gaps 22

[illegible]

RESULT 6

ID	P87295	PRELIMINARY:	PRT:	860 AA.
AC	P87295:			
DT	01-JUN-1998 (TREMBLrel, 06, Created)			
DT	01-JUN-1998 (TREMBLrel, 06, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel, 17, Last annotation update)			
DE	HYPOTHEtical 99.0 KDA PROTEIN Cl6A10.03C IN CHROMOSOME I.			
GN	SPAC16A10.03C.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-972;			
RA	Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;			
RL	Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-1- SIMILARITY: SOME, TO YEAST ENDI.			
DR	EMBL; Z97185; CAB09996.3; -.			
DR	Interpro; IPR001841; Znf_ring.			
DR	SMART; SM00184; RING; 1.			
KW	Hypothetical protein; Transmembrane.			
FT	TRANSMEM 145 165	POTENTIAL.		
FT	TRANSMEM 261 281	POTENTIAL.		
FT	TRANSMEM 431 451	POTENTIAL.		
FT	TRANSMEM 564 584	POTENTIAL.		
SQ	SEQUENCE 860 AA; 99031 MW; 52ABDC9ACCET72CCD CAC64;			

Query Match	4.7%	Score 134	DB 3	Length 860
Best Local Similarity	18.4%	Pred. No. 1.6		
Matches 118; Conservative	85;	Mismatches 234;	Indels 206;	Gaps 19

[illegible]

0

Db 668 -DSLIOIISDPEVTKL-----SETYSESDALHYLKFEVERESITNKY 710
QY 526 LDLETLATSKSSE---DIRLLIQMTEDGAKPEVLAGLLK 565
Db 711 EDLYKILRACFMQFRIPQIOWHVLIVKDGTLNCFELKPLLK 753

RESULT 7
091112 PRELIMINARY; PRT; 321 AA.
AC 091112;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN PA2293.
GN PA2293.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Gardner R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yan Y.,
Brody L.L., Collier S.N., Folger K.R., Kas A., Lardy K., Lim R.M.,
Smith K.A., Spencer D.H., Wong K.E.-S., Wu Z., Paulsen I.T.,
Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen.";
RT Nature 406:959-964(2000).
RL EMBL: AE004653; AAC05681.1; -
DR InterPro: IPR004155; HEAT_PBS.
DR InterPro: IPR000357; HEAT_repeat.
DR Pfam: PF03130; HEAT_PBS; 2.
DR PROSITE: PS50077; HEAT_REPEAT; 1.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 321 AA; 34028 MW; 20A07E919333DB0B CRC64;

Query Match 4.7%; Score 133; DB 16; Length 321;
Best Local Similarity 25.5%; Pred. No. 0.45;
Matches 76; Conservative 31; Mismatches 137; Indels 54; Gaps 10;

QY 70 LKOSHSDPQTRKSTIIGAGLAGSSBALDVLSQAME-TADPQQQLVLVSAGHIGKTS 128
Db 14 LSPRLDADGVRRLALIELADELPALPLVLAALRGDDP-----GVGGEARLL 65
QY 129 D-----DLFLKALASPVVIRLEAAVRLANLKNKTV-----IDHLSPFHKLPPE 173
Db 66 EAMEDAVVALCALADPVAADAASGLGELKEPACGRLLPWGHDAV-----R 120
QY 174 IQCLSAIFLRETEESDAYIRDLAAKSAIRSATALQIGEYOQKRFLLRLMLTS-A 232
Db 121 ASVLRAIRLRELTL--EEBAVALAALAGDPAVAREAVAVGMLRHQALAEKLASADV 178
QY 233 SPDDQEAIVLALGKLKDGOSYINIKQLOKPDVDVTLAAQALIALGKEDALPVYIK-- 290
Db 179 DPEVRRRAATGALISREATVLPALCALADAOQVREAAATTLGKLGREBAGEPLTKALA 238
QY 291 -----QALEERPALVYALRHLPSE-----IGIPALPFIPLKTNSEKLVALLLEL 338
Db 239 DDYWOVRLRARALGRLRHRPARPALEALLGHPI-----GNLRKEAVALALGEL 286

RESULT 8
084133 PRELIMINARY; PRT; 1138 AA.
AC 084133;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE (POSSIBLE TRANSMEMBRANE PROTEIN).
GN CT131.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OC NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL: AE001286; AAC67722.1; -
KW Transmembrane; Complete proteome.
SQ SEQUENCE 1138 AA; 126751 MW; 583015347FE926A5 CRC64;

Query Match 4.7%; Score 133; DB 16; Length 1138;
Best Local Similarity 19.9%; Pred. No. 2.8;
Matches 158; Conservative 119; Mismatches 240; Indels 278; Gaps 41;

QY 8 LFG-LILCSPLISLVA---KEPESVGHKILYISTQSQOALATYLEALDA--YG----- 55
Db 9 IFQYVLVAGVPLALLALPKFSSSESKYLFELSVNKGTLGLOPEIDHLHSMFSQYAKK 68
QY 56 -----DHDFFVLKRGEDYLKQSIHSSDPQ-----TRKSTIIGAGLAGSS 95
Db 69 IRIKIGDSDEIFAAEKI-----LYKGSPLRLLIRFPKALTLQWLSQIDBSLSMNS 121
QY 96 EAL-----DVLQAMETADPLQOL--LVLSAVSG-----HLGKTSDDLKRALA--- 137
Db 122 PSYHLDPGVILSKIERSDITSELGISTMKTNGSTLSVGFYKKAEDLLRLAKREN 181
QY 138 -----SPYPIVRL-----AAVRLNKNKTKYIDHLSPFHKLPPE 172
Db 182 DVGSVAVEGALSPNFEVLNVELSVPASLFRKLPLASLDRLISTEMLINTAKAHQEKD 241
QY 173 EIQLSAIFLRETEESDA---YIRDLAAKSAIRSATALQIGEYOQKRFLLRLN- 227
Db 242 -----STLLITLTAEGNQSIAKLKRGYVDRALFLITOGGASSVLLD-----PTTTSR 286
QY 228 LILSASPD-----QEAILY-----ALGKLKQOSYINIKQLOK-----PDVDVTLAA 271
Db 287 ILSELSPLDPPIRSQEAYLFISEAKLPLSISKWSASDFSLOANLPQISVDPDPNLSIRT 346
QY 272 AQALIALGKEDALPVYIKQ-----ALEERPRA-----L 300
Db 347 ENTKIST-RKSDHLTVIRSSSSAALIGASPSYIHSITSMNKHVAEFHIQHSILPHTYL 405
QY 301 YALNHLRSEIGIPALPFL---KTKNSEAKLNVAL--ALLEGCDTP----- 343
Db 406 RALLPKPIELNTPLEVRYTLDIKGKYRNTQSCALLDNPRLSCNLSGPYALQFSG 465
QY 344 -----KLEFYTERLVQHYNETLALSFSKGTLDONMKRVNIYQDQDERLLST--- 396
Db 466 EGAVTTLSEKWKERLA-DHFLQIQAI-PS-GKMHSQKHVFF-----PKLSGKLIAGDNEI 517
QY 397 -----RGLEE-----QILTF--LFLRPKEAVLPCTYKL----- 422
Db 518 FTHGKFRASEEMKPSNSSLVYGTLSLPDLISPFAPLOMTRYSPSLHSDGKALIK 577
QY 423 -----LASQQTALATTAFSLSHTSQOALDLFQAA---KLPGEPITIR- 463
Db 578 GNVKLFITDPESPFLIEFKILIPDIVISSLDPSAPMTADNISVQASGELLQLPVDRILRL 637
QY 464 AYADLAL-----YNLTKDPEKKRSLHDYAKKLIQETLL-----FVDTE-----NOR 504
Db 638 QHKDLSRYIGTSEASFOILYSPDKKEIV-DISSRFKTDALTGDFRVNKKELSLTEK 696

OY 505 PHSWMLRYOYTYPE-----SRKIMLDILENTLATS--EDIRL 544
 DB 697 THGS---LQWEISPERYSSFEKASPCSCILHPTTIRLDLSCIDKGTGYISL 753
 OY 545 IOLMTEGDAKNPVL 559
 DB 754 SOGIEGKLSSTPLV 768
 RESULT 9
 OYV455 PRELIMINARY; PRT; 514 AA.
 AC OYV455;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE KARYOPHERIN-ALPHA3 PROTEIN.
 GN KARYOPHERIN-ALPHA-3 OR CG9423.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Adill J.F., Agbayani A., An H.-J., Andrews-Piankocch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokva D., Botchan D.A., Bouck J., Brokstein P., Brothier P.,
 Burtis K.C., Busan D.A., Butler H., Caden E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jaitai M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Mason D.A., Goldfarb D.S.;
 RT "Drosophila melanogaster karyopherin alpha 3";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 DR Dockendorff T.C., Tang Z., Jongsens T.A.;

RT "Interaction cloning and characterization of karyopherin alpha 3 from
 RT Drosophila";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003683; AAF54408.1; -
 DR EMBL: AF152928; AAD37442.1; -
 DR HSSP: 002821; 1BK5
 DR FLYBase: FBgn0027338; karyopherin-alpha-3.
 DR InterPro: IPR000225; Armadillo.
 DR InterPro: IPR002652; IBH.
 DR Pfam: PF00514; Armadillo_seg; 8.
 DR Pfam: PF01749; IBH; 1.
 DR SMART: SM00185; ARM; 6.
 DR PROSITE: PS50176; ARM_REPEAT; 2.
 DR SEQUENCE 514 AA; 56990 MW; 0F500AC4EA90683A CRC64;
 SQ
 Query Match 4.7%; Score 132.5; DB 5; Length 514;
 Best Local Similarity 21.2%; Pred. No. 0.96; Indels 137; Gaps 19;
 Matches 98; Conservative 64; Mismatches 164;
 OY 95 SEALDVLISQAMETAD--PLQDLVLVSAVSGHIGKTSDDLFLKALASPPVIREAAVRL 151
 DB 62 SSSIDLKTKAKAAADATKPEQDLAAVQA-----ARKLL 94
 OY 152 ANLKTQVYIDHLSFTIHLPEIOCLSAIFLRTESDAYIRDLAAKSAISATRL 211
 DB 95 SSDKNPPIINDLIQSDI--LPILVECLKQHNHTMLFEAAWL-----TNLSGTS 143
 OY 212 QIGEVQOKRFLPTLNLTSASPODEALIVLAKG-----LRKQGYVNNKKQLQ--- 261
 DB 144 QTNVVAAGAVPLFIQLNSAPNVCQAVNALGNIIGDGLIRFVIRKGVQPLSFI 203
 OY 262 KPDVDVTLA--AAQALALGKEEDALPVYIKQALEERRALVALRH----- 305
 DB 204 KPDIPTFLRVNTVYVIMLCRNKDPAP--PTATHEILPALNVLIHRDNTLVDTWAI 261
 OY 306 -----LPSEIG-IPALPIFLKTNSEAKLV-----ALALLEGCDT----- 342
 DB 262 SYLTGNGDOIQMTVESSGVKPLPL--GNSEVKVQTAALRAVGNIVTGSDEQTVL 318
 OY 343 -----PKLEYTERLVOPHNETALSFSGKRTQGNKRVAIYIP----- 383
 DB 319 NYDALSTPFGILSHPKKEIKRE-----AVWFLSITTAGNSQVQAVNVGLPKRIENL 372
 OY 384 -----QDPERERRLS--TRGLEEOILTFLEPKREAVLPCLYKLASQTLATTAIS 436
 DB 373 SKGEQQTQKEAAMASINTISGNEQVFTL---KEGVIPFDLSCDPTQVINVVL 428
 OY 437 FLSH-----TSHQEL-DLFLQAKLGEPTIRAVADLAITNL 473
 DB 429 GLNNMLKVAADSHVAVANCICEGCLAKIERIQSHENVEIYKL 471
 RESULT 10
 OYU1K7
 ID OYU1K7 PRELIMINARY; PRT; 948 AA.
 AC OYU1K7;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE ELKS.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakata T.;
 RT "Fusion of a Novel Gene, ELKS, to c-ret in a Papillary Thyroid
 RT Carcinoma";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB013617; BAA8763.1; -

DR InterPro: IPR002017; Spectrin.
SQ SEQUENCE 948 AA; 108792 MW; 344297FDFC9F7602 CRC64;

Query Match 4.7%; Score 132.5; DB 4; Length 948;
Best Local Similarity 20.1%; Pred. No. 2.3;
Matches 116; Conservative 106; Mismatches 204; Indels 151; Gaps 28;

OY 35 ISTOSTQALATYLEDAYGDHDFVLRKIGEDYLKSHSSDP-----QTRKSTIIIGA 89
DB 414 LSTEEEREE---MKOMEYVRSHSKFKNNKIGQ--VKQESRKDTELLALQTKLETLLNQ 467
OY 90 GLAGSSALDVLSCAMETADPLQOLVLVSAVSGHIGKTSDDLFLKALASPYVIRLEAY 149
DB 468 -FDSQKHIEIKESLSLAKE--QRAAIL-----QTEVDAL-----RLRLKEKE 507
OY 150 RLANKTKYID-----HLHSF-----THKLPETIOCSAAI----- 181
DB 508 TMLNKKTKROIODMAEEKGTAGEIHDLKMDLVKERVNVYLQKKIKTENLQEQLRDKEROMS 567
OY 182 -----FLRETEESDAYIRDL---LAAKSATSATALQIGEVQOKR-----FLPTLR 226
DB 568 SLKERVKSLQADVTNTPTALTTEELAEKERTIERLEKQDRDEREKQOEIDVYKKDLK 627
OY 227 NL-----LTSASPDQDEAIIYALGKLKDGOSYVNIKKOLQKPDVDTLAAQAALIALGKE 281
DB 628 DLKRVKSLQGDLSKEKASLLDLKEHASSLASSGLKDSRLKTEIALEQ-----KK 679
OY 282 EDALPV---IKK---QALEER--PRALYALRHPSEIGIPALPIFLTKNSEAKLNVAL 333
DB 680 EECLKMESOLKKAHEALAEARASPEMSDRIOHLEREI-----TRYKDESSKAQAEV-- 730
OY 334 ALLEGGDTPKLIEYTERLVQPHYNE-----TLALSFGSGRTLQNMKRVNIIVPODQ 387
DB 731 -----DRLEETLKE--VENEKNDKDKIALESLSTSQVQDNKKVANLKHKEVE 779
OY 388 ERE--RLSTTRGLEEOILFELFLPREAYLPCYKL---LASQOTOLATTAISFLSHTS 442
DB 780 KKKSAQMLEARREDNLNDSQOLQVEELMAHEKVAQOELESMAKAKLSIQOSLAERET 839
OY 443 H-----QEAIDLFOAKKLPPEPIIRAY-----ADLAIYNLTKDEPKKRSLDYAKKIQE 493
DB 840 HLTNLRAERKRKHLEVELEMKQEAALLAISEKDANIALLELSS-----SKKTKQE 888
OY 494 TILFVDTENORHPSPMYRLRYQVPESTRKILMILE 530
DB 889 EVALKREKDR--LVQOLKQOT--QNMKMLADNYE 920

RESULT 11
O9HKNG PRELIMINARY; PRT; 1200 AA.

AC O9HKNG;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOHETICAL PROTEIN TA0561.
GN TA0561.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN-DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Repp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
Mewes H.-W., Frisman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
acidophilum";
RL Nature 407:508-513(2000).
DR EMBL; AL45064; CAC11701.1; -
DR InterPro; IPR000547; Clathrin_repeat.

DR InterPro: IPR001623; DnaJ_N.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 4.
DR SMART: SM00289; CLH; 1.
DR PROSITE: PS00636; DNAJ_1; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1200 AA; 138346 MW; B056E2030B945418 CRC64;

Query Match 4.7%; Score 132.5; DB 17; Length 1200;
Best Local Similarity 19.8%; Pred. No. 3.3;
Matches 115; Conservative 88; Mismatches 209; Indels 169; Gaps 25;

OY 98 LDVLSQAMETADP-----LQOLVLVSAVSGHIGKTSDDLFLKALASPYVIRLEAY 149
DB 99 LDVI--REVTSDFPCEVENIVNLVISA--DHLEKYTFVAFILSKGCRVD--DRIAAYYA 153
OY 150 -RLANKTKYIDHLHSFIHKLPEETIOCSAIFLR--LETESDAYIRDLA----- 199
DB 154 ERDRDDSTTDII-----IDNYEKHNDGRAVISFLRHVLAYEYSERYITKLECGRVED 207
OY 200 -----AKSATSATALQIGEVQOKRFLPTLRNLTSASPDQEA- 240
DB 208 SESYIYAVELAEPPERVENADSSAILARTALDLGKEKARQIAE--RGILKLPDSEDLKLM 266
OY 241 ---LYALGKLKDGGSY-----NIKKOLQK 262
DB 267 ARSLYALGRINESLDFYDVCNHYPERNEANYEMMDILYNGRTKEVATETLNVHRESFR 326
OY 263 PD-----VDVTLAAQAALIALGKEDALPVTKQALEERP-----RALYALRH 305
DB 327 PEDYIRMSGLIKDSGDVGSVNLLEKALKEPPNDIIDIKAYAAAKKEIGNVGEALQAVON 386
OY 306 L-----PSEIGIPALPIFLTKNSEAKLN-----ALALLEGC--DT 342
DB 387 LRIKPDDESLKFNVDYFFGRNEDILNYESLQDALLKERYGMAAASYVMGYMDA 446
OY 343 KLLYEYTERLVQPHYNETLALSFGSGRTLQNMKRVNIIVPODERERLSTTR--GLE 400
DB 447 VSMTRDHEILLDDPFVDSILFSVAKK-----EEVQMLISGLENYTSRLSVDRFLGIE 500
OY 401 ---EQILFELR-----LPKEAYLPCYKLLASQOTOLATTAISFLSHTSQOELD 448
DB 501 IRGVQYIIDYATKSCSKAMAEVAAVAFRKNHTVDKIKVALSTKCL-----EYYD 552
OY 449 LLFOAKLPGERPIIRAYADLAIYNLTG--DPEKRSLSHYAKKLIQETILFVDTENQORHP 507
DB 553 IMISMGIYSKENVKIFEDHPHYLPAYIDFTTAGLYDEAYKILR-----YEGTKEDP 606
OY 508 SMPY---LRYQVPESTRKILMILETLATSKSSEDRL 544
DB 607 FLDIYEARLMTVM---NRTKDAIKLIRAKESFNSIDRYLL 644

RESULT 12
O30981 PRELIMINARY; PRT; 2554 AA.

AC O30981;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FENGYCIN SYNTHETASE FENE.
GN FENE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN-F29-3;
RA Chen I., Lin G., Shu H., Liu S.;
RT "Analysis of the FenGYCIN Synthetase Gene fene";
RL Thesis (1997), Microbiology and Immunology, Chang-Gung University.

DR EMBL: AF023465; AAB80956.1; -
 DR HSSP: P14687; 1AMU.
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR001242; DUF4.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR Pfam: PF00501; AMP-binding; 2.
 DR Pfam: PF00668; Condensation; 3.
 DR Pfam: PF00550; pp-binding; 2.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00075; ACP_DOMAIN; 2.
 DR PROSITE: PS00455; AMP_BINDING; 2.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.
 DR Phosphopantetheine.
 KW SEQUENCE 2554 AA; 286724 MW; AEF0E71BF77AC0 CRC64;

Query Match 4.7%; Score 132.5; DB 2; Length 2554;
 Best Local Similarity 20.0%; Pred. No. 9.7;
 Matches 105; Conservative 80; Mismatches 186; Indels 155; Gaps 21;

QY 122 GHLGTSDDLFFKALASPPVIRLEAAYRLANKTKYIDHLHSFIHKLPPEIQCLSAAI 181
 DB 1851 GYLGRPD---LTKEFEVNPFPAPGQMYRTGDLARW-LPDGTIEYGRVDDQVKIRG--- 1903
 QY 182 FLRLTEPSDAYIRDLAALKKSAIRSATALQIGEVQOKRFLPTLNLTLSASPOQEAAL 241
 DB 1904 -YRVLGETESALRHIDGVKEAAVLAFTG-QLGTRK-----L 1938
 QY 242 VALGKIDGOSYYNKKLOKEDVDVTLAAQALATLAGEEDALPV---IKQALEER 296
 DB 1939 YAVISVKEGTDEQYRTHLSQ-----MLPGYMPAYIEMALPLTNGKLNKRALEP- 1991
 QY 297 PRALYALNHLSEIGIPALPIFLTKNSEAKLNVALALLEIGCTPKLE----- 347
 DB 1992 PDITSKQTYVPVRNDLEOLAIHQEVLGTQIGIEDSFEELGDSIKALQVARSALRGY 2051
 QY 348 ---YTERLYOVPHNETLALSFSGRT-----IQNKRNIIVYPOPORE 390
 DB 2052 WSLHNSDLFRAPKIKDLSAIVKTERVYIDGOSGAVFWPTPLQHM-----FLSDIKERH 2106
 QY 391 R-----LSTRTGLEEQILTEFLRLPKRAY--LPCITYKLASOKTO-----LATTA 434
 DB 2107 HFNQSVMLFSPDCLSENMLRASLKLAEHHDALRMITYREDSGQGMQINODIHESLYSLR 2166
 QY 435 ISFLS-----HTSHQELDLFLQAALPGEPITRA-----YADLAIYNLTKD--- 476
 DB 2167 ISDLSDSGMDWETSIKEEVANLQOSINLQOGPLHAAMFKTLSDYLFLLTHHILVVDGVS 2226
 QY 477 -----PEKKRSLHDYAK-----KLIOETLLFVDTENOR 504
 DB 2227 WRILLEDSAAVHOASQAOLQLPKPTDSTYOETARRVODVYAOSSSKLIREETYSVSEEEK 2286
 QY 505 ----PHPSPMYLRYQVTPESRTKMLDLLETLANSSKSESDIRLLIQ 546
 DB 2287 AAEPLPY-EMPYMENKNSSE-----ETLRSLTEADNAVLLQ 2322

RESULT 13
 Q44295 PRELIMINARY; PRT; 398 AA.
 AC 044295;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE ORF398 PROTEIN.
 GN ORF398.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-PCC 7120;
 RX MEDLINE=97144534; PubMed=8990301;

RA Frías J.E., Flores E., Herrero A.;
 RT "Nitrate assimilation gene cluster from the heterocyst-forming
 RT cyanobacterium Anabaena sp. Strain PCC 7120.";
 RL J. Bacteriol. 179:477-486(1997).
 DR EMBL: X99708; CAA68039.1; -
 DR InterPro: IPR004155; HEAT_PBS.
 DR Pfam: PF03130; HEAT; 9.
 DR SEQUENCE 398 AA; 43348 MW; A38946D4ID5AF05E CRC64;

Query Match 4.6%; Score 131; DB 2; Length 398;
 Best Local Similarity 22.6%; Pred. No. 0.84;
 Matches 90; Conservative 38; Mismatches 123; Indels 148; Gaps 12;

QY 31 KILYISTQOALATYLEALDAYGDHDFVLRKIGEDYLYKOSIHSSDPQTRKSTLIGAG 90
 DB 65 KVLVNLGNVTINPLDILDEDEDAEDLRWFARILIGE----- 101
 QY 91 LAGSSEALDVSQAMETADPLQOOLVLSAVSGHLCKTSDDLFFKALAS-----PPY 141
 DB 102 -----LQHPREALPVELL-----KTSDEEIKAIASSAIAQMGTLAIP 140
 QY 142 VI-----RLFAAYRLANKTKYIDHLHSFIHKLPPEIQCLSAAIFFRLTEESDA 192
 DB 141 VIVELLAQENTRLAVRSLAYIRHTQTIALPLSVQ-----DTQAS-- 181
 QY 193 YIRDLAALKKSAIRSATALQIGEVQOKRFLPTLNLTLSASPOQEAALYALGKIDGOS 252
 DB 182 -----IRAAAEIALSFHDQVPPFLNALNDLSTVTRTAIQGSFSPDLS 229
 QY 253 YINIKQLOKPR---DVDTYLAQAQALATLAGEEDALPYIKQALEER-----RAL 300
 DB 230 ELNLVAQIQPKLYDNIENCCAAANALAMGDDAAOHLKYLISAHNPITLQLEIRAL 289
 QY 301 YALRHLP-----EIGIPALPIFLTKNS----- 325
 DB 230 VMLESLISLEYIQCAFNOVTTETLMQELVTVYGRQKREKLP-ATAILLELNSPHRAT 348
 QY 326 ---EAKLNVALALLEG---CDPRKLEYYTERLVQPH 357
 DB 349 KHSKSAIALSLGOLGNAAEATESLTMLSTPDELAVRH 387

RESULT 14
 Q9EP71 PRELIMINARY; PRT; 979 AA.
 AC 09EP71;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE NORPEG-LIKE PROTEIN (ANKYCORBIN).
 GN 1700020L1IRK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=C57 BLACK/6;
 RA Peng Y.-F., Mandal K., Sakisaka T., Okabe N., Yamamoto Y.,
 RA Yokoyama S., Mizoguchi W., Shiozaki H., Monden M., Takai Y.;
 RT "Ankycorbin: a novel actin cytoskeleton-associated protein.";
 RT Genes Cells 0:0-0(2000).
 DR EMBL: AF274866; AAG25937.1; -
 DR EMBL: AF202315; AAG24483.1; -
 DR HSSP; P42773; ITHB.

DR MGD; MG1:1922896; 1700020L1IRIK.
 DR InterPro; IPR002110; ANK.
 DR PRINTS; PRO1415; ANKYRIN.
 DR SMART; SM00248; ANK; 6.
 DR PROSITE; PSS0086; ANK_REPEAT; 5.
 DR PROSITE; PSS0297; ANK_REPEAT; 1.
 DR ANK repeat; Repeat.
 KW ANK repeat; Repeat.
 SO SEQUENCE 979 AA; 108851 MW; B2B8C016D80237C4 CRC64;

Query Match 4.6%; Score 130.5; DB 11; Length 979;
 Best Local Similarity 20.5%; Pred. No. 3.3;
 Matches 127; Conservative 89; Mismatches 224; Indels 181; Gaps 28;

QY 32 ILVTSQSTQOALATVLEA-----LDAYGHDFVLRKIGEDYIKOSIHS 76
 DB 336 LLDISSADODDLVLQAAVASITLHNKELQDKLQAKSKPD-----KEARDISFQSFHS 391
 QY 77 SD-----PQTRKSTLIGAGLASSSEALD-----VLSQAMETADPLQQL----- 115
 DB 392 TQTDLAPSPGKASDIPESDAKSSPVHPAGTSTTDNDVILIRQLQSDSLHDLQKRLESSEA 451
 QY 116 -----VLSVSGHIGKTSDDL--LFKALASPY-----PVIRLEAAYPLAN 153
 DB 452 EKKLODELQSORDTLCLINTEISENGSDLSOKLKTQSKYEAMKEVLSVOKOMKIGL 511
 QY 154 LKNTKVIDLHSLFHKPEEIOCSAIFLRLETEES--DAYTRD--LAAKSAIRS 207
 DB 512 LSGESADGYSHLREAPADEIDTLKQD--LQKAVESARNRKERELETETLAEKQAEAT 569
 QY 208 ATALQIGEYQOKREPLPLRN-----LTSASPODEAILYALGKLKDGQSYNYIKOL- 260
 DB 570 KPPAEACEELRSYSYIENNNKKAFLFEKYQQAQGEIM-----KLMD-----TLKSQMP 620
 QY 261 -QKPD--VDYTLAAQALILGKEEDALPYI-----KKQALERPRLALYLRH 305
 DB 621 QEAPDSDGDKKAMNNRMLDELNKOVSSELSQLYREAOQLDLYDRKRSLEDAAEYIHKAH 680
 QY 306 LPSEIGIPALPIFLKTRN--SEAKLNVALALLETGCDTPKLEETE--RLVQPHYMETL 362
 DB 681 -----ERLMHNSNLSRAKSEALSS--EKKSOYSKVLNLTQLKLVQAH----- 722
 QY 363 ALFSKGRTLQNMKNRNIIVPODERERLLSTRTGLEQILTFELRP--KEAVLPCTYK 421
 DB 723 -----KENSVESTENHQLVITTLRTAKEMEKISALTGLHAKAEAVALEK 769
 QY 422 LLSQKQLATTAISLSHTSHOEALDLFQAALPGEPIIRAYADIAIYNLTKDPEKR 481
 DB 770 QLAEEKAASVDAMVPKSSYEKLQASL-----ESEVNALA-----TKLKSVR 811
 QY 482 SLHDYAKKLQOETLLEFDTENORHPSPMLRYQVTPESRKLMLDLLETATSKSEDI 541
 DB 812 -----BREKASHSEVAQVRSVSQAREK--DNIOQLTKAKEQD-V 848
 QY 542 RLIIQLMTEGDANKFPYIAGL 562
 DB 849 TALVQKFORAQEE-----LAGM 865

RESULT 15
 Q9M6P9 PRELIMINARY; PRT; 1316 AA.
 AC Q9M6P9;

DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE DNA REPAIR-RECOMBINATION PROTEIN.
 GN RAD50.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA.
 RX MEDLINE=21097002; PubMed=11169180;
 RA Gallego M.E., Jeanneau M., Granier F., Bouchez D., Bechtold N.,
 RA White C.I.;
 RT "Disruption of the Arabidopsis RAD50 gene leads to plant sterility and
 RT RMS sensitivity";
 RL Plant J. 25:31-41(2001).
 DR EMBL; AF168748; AAF36810.1; -
 DR InterPro; IPR003439; ABC_transport.
 DR InterPro; IPR001687; ATP_GTP_A.
 SO SEQUENCE 1316 AA; 152814 MW; 89DC4F6BCA39B0E8 CRC64;

Query Match 4.6%; Score 129.5; DB 10; Length 1316;
 Best Local Similarity 22.0%; Pred. No. 5.9;
 Matches 127; Conservative 93; Mismatches 229; Indels 129; Gaps 25;

QY 62 LRKIGEDYIKOSI-----HSSDPQTRKSTI-----IGAGLASSSEALDVLSQA 104
 DB 745 LRATVEEYSKLTITITPLAEKTELOEHELGQKSEALDVLIGISAQIKADKDSIEALVQ 804
 QY 105 METADPLQQLLVLSAVSGHIGKTSDDLFRALASPYVIRLEAY--RLANKNTKVIDHL 163
 DB 805 LENADRIFQETIV-----SYQKQIEDLEYKLDPRGLGVTKTMEETSELSQSSK--DKL 856
 QY 164 HSFHFKPE-----EIQCSAIF--LRLETESDAVIRPLAKSAIRSATALQGE 215
 DB 857 HGELEKLRDQIYMERISCIQARWHAVEREKAANILRLVTAEDLERLA----- 909
 QY 216 YQKRFPLTRNLNTSA--SPQDEAIL--YALGKLKDGQSYNYI--KKQLOKPDVYT 268
 DB 910 -EEKSQDLQVYKYLTEALGPLSKREKQLDSYNNMKIRRNQVEYELAEKKRNYQGEVAL 968
 QY 269 LAAQALIA--IGKEDALPVIKQALEERPRLALYLRHLPSEIGIPALPIFLKTRNS 325
 DB 969 LKASYKINEYHDLKGERLIDQEKORLSDS-----QLOSC 1004
 QY 326 EAKLNVALAL--ELGCDTPKLEETEYTELQVPHYNETLALSFSGRTLQNMK-- 377
 DB 1005 EARNELAGELNRNKLDRNDQLRNIEDNL--NRTTKAKVEELRLETESLEQITLN 1061
 QY 378 -----VNIIVPODERERLLS-----TTRGLEQILTFELRPKEAYLPCTYKLA 424
 DB 1062 IGGIAVAEAEIVKILREBERLLSELNCRGVSYESSISNNRVE--KQAOYKIDIKRH 1120
 QY 425 SOKTOLATTAISLSHTSHOEALD--LRFQAALPG--EPIIRAY-----ADLAIYNL 473
 DB 1121 DQLIQLTKTEKANKDLDRIYALDKALMRFHTMMEELINKIIRELMQOTYRGQMDYIRI 1180
 QY 474 TKDPEKRSLSHDYAKKLQOETLLEFDTENORHPSPMLRYQVTPESRKLMLDLLET 533
 DB 1181 HSDSE--GAGTRSYSVKYLMOY--GDTE-----LEWRGCSAGOKVLASLIIRLA 1226
 QY 534 TSKS--SEDIRLIIQLMTEGDANKFPYIAGLIIKIVE 568
 DB 1227 LAETFCILNCGILALDEPTTILNDGPNSESLAGALLIRIME 1264

Search completed: August 20, 2002, 04:23:45
 Job time: 809 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:19:11 ; Search time 17.02 Seconds
(without alignments)
815.143 Million cell updates/sec

Title: US-09-662-812-2

Perfect score: 568
Sequence: 1 MGLFHLTLGLLCSLPISL.....TEGDKNPPVLAGLIKRIKIVE 568

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/pcodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_6/pcodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_6/pcodata/2/1aa/5A.COMB.pep:*
- 4: /cgn2_6/pcodata/2/1aa/5B.COMB.pep:*
- 5: /cgn2_6/pcodata/2/1aa/PCOTUS.COMB.pep:*
- 6: /cgn2_6/pcodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	1.2	11	1	US-07-657-769B-43 Sequence 43, Appl
2	7	1.2	11	1	US-07-789-184-104 Sequence 104, App
3	7	1.2	11	1	US-07-789-184-215 Sequence 215, App
4	7	1.2	11	1	US-07-789-184-218 Sequence 218, App
5	7	1.2	11	1	US-08-475-263-104 Sequence 104, App
6	7	1.2	11	1	US-08-475-263-215 Sequence 215, App
7	7	1.2	11	1	US-08-475-263-218 Sequence 218, App
8	7	1.2	11	1	US-08-485-886-104 Sequence 104, App
9	7	1.2	11	1	US-08-485-886-215 Sequence 215, App
10	7	1.2	11	1	US-08-485-886-218 Sequence 218, App
11	7	1.2	11	2	US-08-477-362-104 Sequence 104, App
12	7	1.2	11	2	US-08-477-362-215 Sequence 215, App
13	7	1.2	11	2	US-08-477-362-218 Sequence 218, App
14	7	1.2	11	2	US-08-477-134-104 Sequence 104, App
15	7	1.2	11	2	US-08-477-134-215 Sequence 215, App
16	7	1.2	11	2	US-08-477-134-218 Sequence 218, App
17	7	1.2	11	3	US-08-473-489A-104 Sequence 104, App
18	7	1.2	11	3	US-08-473-489A-215 Sequence 215, App
19	7	1.2	11	3	US-08-473-489A-218 Sequence 218, App
20	7	1.2	11	3	US-08-485-695-104 Sequence 104, App
21	7	1.2	11	3	US-08-485-695-215 Sequence 215, App
22	7	1.2	11	3	US-08-485-695-218 Sequence 218, App
23	7	1.2	11	4	US-08-018-760-104 Sequence 104, App
24	7	1.2	11	4	US-08-018-760-215 Sequence 215, App
25	7	1.2	11	4	US-08-018-760-218 Sequence 218, App
26	7	1.2	151	4	US-08-905-223-436 Sequence 436, App
27	7	1.2	186	4	US-08-936-165A-534 Sequence 534, App

28	7	1.2	226	3	US-08-651-136C-16	Sequence 16, Appl
29	7	1.2	247	1	US-08-446-083-5	Sequence 5, Appl
30	7	1.2	293	3	US-08-651-136C-20	Sequence 20, Appl
31	7	1.2	298	3	US-08-651-136C-18	Sequence 18, Appl
32	7	1.2	322	1	US-08-118-270-36	Sequence 36, Appl
33	7	1.2	322	5	PCIT-US93-08528-36	Sequence 36, Appl
34	7	1.2	331	4	US-08-849-751-4	Sequence 4, Appl
35	7	1.2	331	4	US-09-413-231-4	Sequence 4, Appl
36	7	1.2	331	4	US-09-478-816-4	Sequence 4, Appl
37	7	1.2	335	2	US-08-505-218-4	Sequence 4, Appl
38	7	1.2	367	4	US-08-888-429A-13	Sequence 13, Appl
39	7	1.2	425	1	US-07-657-769B-69	Sequence 69, Appl
40	7	1.2	425	1	US-08-097-938-7	Sequence 7, Appl
41	7	1.2	425	1	US-08-313-553-13	Sequence 13, Appl
42	7	1.2	425	1	US-07-789-184-220	Sequence 220, App
43	7	1.2	425	1	US-08-476-000-7	Sequence 7, Appl
44	7	1.2	425	1	US-08-475-263-220	Sequence 220, App
45	7	1.2	425	1	US-08-472-840-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-07-657-769B-43
; Sequence 43, Application US/07657769B
; Patent No. 5256766
; GENERAL INFORMATION:
; APPLICANT: COUGHLIN, SHAUN R.
; TITLE OF INVENTION: RECOMBINANT RECEPTOR AND RELATED
; NUMBER OF INVENTIONS: PHARMACEUTICALS
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IRELL & MANELLA
; STREET: 545 MIDDLEFIELD ROAD, SUITE 200
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/657,769B
; FILING DATE: 19910219
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2000-0502.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-327-7250
; TELEFAX: 415-327-2951
; TELEX: 706141
; INFORMATION FOR SEQ. ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-657-769B-43

Query Match 1.2%; Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 436 SFLSHTS 442
Db 3 SFLSHTS 9

*Issued
Patents*

RESULT 2
US-07-789-184-104
Sequence 104, Application US/07789184
Patent No. 5688768
GENERAL INFORMATION:
APPLICANT: COUGHLIN, SHAUN R.
ATTORNEY/AGENT INFORMATION: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
TITLE OF INVENTION: RELATED PHARMACEUTICALS
NUMBER OF SEQUENCES: 223
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789,184
FILING DATE: 19911107
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20502.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-789-184-104

Query Match 1.2%; Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
|||||||
DB 3 SFLSHTS 9

RESULT 3
US-07-789-184-215
Sequence 215, Application US/07789184
Patent No. 5688768
GENERAL INFORMATION:
APPLICANT: COUGHLIN, SHAUN R.
ATTORNEY/AGENT INFORMATION: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
TITLE OF INVENTION: RELATED PHARMACEUTICALS
NUMBER OF SEQUENCES: 223
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789,184
FILING DATE: 19911107
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20502.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-789-184-218

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789,184
FILING DATE: 19911107
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20502.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 215:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-789-184-215

Query Match 1.2%; Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
|||||||
DB 3 SFLSHTS 9

RESULT 4
US-07-789-184-218
Sequence 218, Application US/07789184
Patent No. 5688768
GENERAL INFORMATION:
APPLICANT: COUGHLIN, SHAUN R.
ATTORNEY/AGENT INFORMATION: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
TITLE OF INVENTION: RELATED PHARMACEUTICALS
NUMBER OF SEQUENCES: 223
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789,184
FILING DATE: 19911107
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20502.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-789-184-218

Query Match 1.2%; Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
Db 3 SFLSHTS 9

RESULT 5
US-08-475-263-104
; Sequence 104, Application US/08475263
; Patent No. 5759994
; GENERAL INFORMATION:
; APPLICANT: COUGHLIN, SHAUN R.
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,263
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELETYPE: 90-4030
; INFORMATION FOR SEQ. ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-475-263-104

Query Match 1.2%; Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
Db 3 SFLSHTS 9

RESULT 6
US-08-475-263-215
; Sequence 215, Application US/08475263
; Patent No. 5759994
; GENERAL INFORMATION:
; APPLICANT: COUGHLIN, SHAUN R.
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
; NUMBER OF SEQUENCES: 223

CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Ave., NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,263
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20502.03
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELETYPE: 90-4030
INFORMATION FOR SEQ. ID NO: 215:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-475-263-215

Query Match 1.2%; Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
Db 3 SFLSHTS 9

RESULT 7
US-08-475-263-218
; Sequence 218, Application US/08475263
; Patent No. 5759994
; GENERAL INFORMATION:
; APPLICANT: COUGHLIN, SHAUN R.
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,263
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 22000-20502.03
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-485-886-218

Query Match 1.2%: Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
|||||||
Db 3 SFLSHTS 9

RESULT 8
US-08-485-886-104
Sequence 104, Application US/08485886
Patent No. 5798248
GENERAL INFORMATION:
APPLICANT: COUGHLIN, SHAUN R.
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
TITLE OF INVENTION: RELATED PHARMACEUTICALS
NUMBER OF SEQUENCES: 223
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,886
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/789,184
FILING DATE: 07-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20502.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-485-886-104

Query Match 1.2%: Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 436 SFLSHTS 442
|||||||

Db 3 SFLSHTS 9

RESULT 9
US-08-485-886-215
Sequence 215, Application US/08485886
Patent No. 5798248
GENERAL INFORMATION:
APPLICANT: COUGHLIN, SHAUN R.
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
TITLE OF INVENTION: RELATED PHARMACEUTICALS
NUMBER OF SEQUENCES: 223
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,886
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/789,184
FILING DATE: 07-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20502.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 215:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-485-886-215

Query Match 1.2%: Score 7; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
|||||||
Db 3 SFLSHTS 9

RESULT 10
US-08-485-886-218
Sequence 218, Application US/08485886
Patent No. 5798248
GENERAL INFORMATION:
APPLICANT: COUGHLIN, SHAUN R.
APPLICANT: SCARBOROUGH, ROBERT M.
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
TITLE OF INVENTION: RELATED PHARMACEUTICALS
NUMBER OF SEQUENCES: 223
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California

COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/485,886
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/789,184
 FILING DATE: 07-NOV-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: MURASHIGE, KATE H.
 REGISTRATION NUMBER: 29,959
 REFERENCE/DOCKET NUMBER: 22000-20502.20
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 813-5600
 TELEFAX: (415) 494-0792
 TELEX: 34-0154
 INFORMATION FOR SEQ ID NO: 218:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 11 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-485-886-218

Query Match 1.2%; Score 7; DB 1; Length 11;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 436 SFLSHTS 442
 Db 3 SFLSHTS 9

RESULT 11
 US-08-477-362-104
 Sequence 104, Application US/08477362
 Patent No. 5849507
 GENERAL INFORMATION:
 APPLICANT: COUGHLIN, SHAUN R.
 APPLICANT: SCARBOROUGH, ROBERT M.
 TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
 NUMBER OF SEQUENCES: 223
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 Page Mill Road
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/477,362
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/789,184
 FILING DATE: 07-NOV-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: MURASHIGE, KATE H.
 REGISTRATION NUMBER: 29,959
 REFERENCE/DOCKET NUMBER: 22000-20502.20

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 813-5600
 TELEFAX: (415) 494-0792
 TELEX: 34-0154
 INFORMATION FOR SEQ ID NO: 104:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 11 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-477-362-104

Query Match 1.2%; Score 7; DB 2; Length 11;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 436 SFLSHTS 442
 Db 3 SFLSHTS 9

RESULT 12
 US-08-477-362-215
 Sequence 215, Application US/08477362
 Patent No. 5849507
 GENERAL INFORMATION:
 APPLICANT: COUGHLIN, SHAUN R.
 APPLICANT: SCARBOROUGH, ROBERT M.
 TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
 NUMBER OF SEQUENCES: 223
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORRISON & FOERSTER
 STREET: 755 Page Mill Road
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/477,362
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/789,184
 FILING DATE: 07-NOV-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: MURASHIGE, KATE H.
 REGISTRATION NUMBER: 29,959
 REFERENCE/DOCKET NUMBER: 22000-20502.20
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 813-5600
 TELEFAX: (415) 494-0792
 TELEX: 34-0154
 INFORMATION FOR SEQ ID NO: 215:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 11 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-477-362-215

Query Match 1.2%; Score 7; DB 2; Length 11;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 436 SFLSHTS 442

Db 3 SFLSHTS 9

RESULT 13

US-08-477-362-218
; Sequence 218, Application US/08477362
; Patent No. 5849507
; GENERAL INFORMATION:
; APPLICANT: COUGHLIN, SHAUN R.
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
; TITLE OF INVENTION: RELATED PHARMACEUTICALS
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,362
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,184
; FILING DATE: 07-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 22000-20502.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 34-0154
; INFORMATION FOR SEQ ID NO: 218:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-477-362-218

Query Match 1.2%; Score 7; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
Db 3 SFLSHTS 9

RESULT 14

US-08-477-134-104
; Sequence 104, Application US/08477134
; Patent No. 5856448
; GENERAL INFORMATION:
; APPLICANT: COUGHLIN, SHAUN R.
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
; TITLE OF INVENTION: RELATED PHARMACEUTICALS
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto

STATE: California
COUNTRY: USA
ZIP: 94304-1018

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,134
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,184
; FILING DATE: 07-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 22000-20502.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 34-0154
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-477-134-104

Query Match 1.2%; Score 7; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442
Db 3 SFLSHTS 9

RESULT 15

US-08-477-134-215
; Sequence 215, Application US/08477134
; Patent No. 5856448
; GENERAL INFORMATION:
; APPLICANT: COUGHLIN, SHAUN R.
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND
; TITLE OF INVENTION: RELATED PHARMACEUTICALS
; NUMBER OF SEQUENCES: 223
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,134
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/789,184
; FILING DATE: 07-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959

```

; REFERENCE/DOCKET NUMBER: 22000-20502.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 34-0154
; INFORMATION FOR SEQ ID NO: 215:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-477-134-215
    
```

```

Query Match          1.2%; Score 7; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    
```

```

OY 436 SFLSHTS 442
   |||||
Db 3 SFLSHTS 9
    
```

Search completed: August 20, 2002, 04:25:56
 Job time: 405 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 01:09:15 ; Search time 1588.14 Seconds
(without alignments)
16206.816 Million cell updates/sec

Title: US-09-662-812-1

Perfect score: 1907

Sequence: 1 gtgagctgatttgaaaaag.....ctcatcaagagagctgtgaaa 1907

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estp1:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	1.1	458	11 AK019607	AK019607 Mus muscu
2	21	1.1	443	9 AM098474	AM098474 ga03g06.x
3	21	1.1	534	9 AM680975	AM680975 WS1_9_H08
4	21	1.1	582	9 AM564895	AM564895 LG1_312_D
5	21	1.1	909	12 A2538423	A2538423 ENTFT26TF
6	20	1.0	216	10 BG959111	BG959111 PM4-CT080
7	20	1.0	224	9 AA669611	AA669611 ac19e05.s
8	20	1.0	326	10 BB649446	BB649446 UI-R-BH2.
9	20	1.0	370	9 AV855875	AV855875 AV855875
10	20	1.0	393	12 AQ124704	AQ124704 HS_2266.A
11	20	1.0	433	9 A1159433	A1159433 v279b05.f
12	20	1.0	452	10 BM409407	BM409407 EST583734
13	20	1.0	470	10 BJ073983	BJ073983 BJ073983
14	20	1.0	491	10 BJ729956	BJ729956 603350062
15	20	1.0	496	10 BM409927	BM409927 EST584254
16	20	1.0	510	10 BG098375	BG098375 EST462894
17	20	1.0	563	10 BE436245	BE436245 EST407323

C 18	20	1.0	567	10 BE433437	BE433437 EST39966
C 19	20	1.0	581	12 AQ276990	AQ276990 CTRB1-E1-
C 20	20	1.0	590	9 AV852345	AV852345 AV852345
C 21	20	1.0	593	9 AV680699	AV680699 AV680699
C 22	20	1.0	621	9 A1486589	A1486589 EST244910
C 23	20	1.0	627	10 B1730573	B1730573 603350953
C 24	20	1.0	645	10 BM387094	BM387094 UI-R-CN1-
C 25	20	1.0	660	9 A1898005	A1898005 EST267448
C 26	20	1.0	673	9 BB649694	BB649694 BB649694
C 27	20	1.0	687	10 B17289662	B17289662 UI-R-DK0-
C 28	20	1.0	705	9 A1898469	A1898469 EST267912
C 29	20	1.0	713	10 BM411178	BM411178 EST585505
C 30	20	1.0	714	10 BG967518	BG967518 602834090
C 31	20	1.0	744	10 BM412698	BM412698 EST587025
C 32	20	1.0	759	9 AU080658	AU080658 AU080658
C 33	20	1.0	815	10 BG975641	BG975641 602845351
C 34	20	1.0	835	10 BF784222	BF784222 602108086
C 35	20	1.0	963	12 CNS03MW3	CNS03MW3 Tetradon
C 36	20	1.0	972	10 BG295852	BG295852 602393359
C 37	19	1.0	101	9 BE076598	BE076598 CM1-BT059
C 38	19	1.0	133	9 AA706483	AA706483 ag95d11.s
C 39	19	1.0	290	9 BR391458	BR391458 BR391458
C 40	19	1.0	326	9 AA09747	AA09747 OK88B09.s
C 41	19	1.0	350	9 AA961192	AA961192 CO52601.s
C 42	19	1.0	357	9 A1451953	A1451953 mp75g06.x
C 43	19	1.0	373	10 BG930973	BG930973 f33-1134
C 44	19	1.0	380	10 BF018004	BF018004 uw96b04.x
C 45	19	1.0	380	10 BF018281	BF018281 uw97a10.x

ALIGNMENTS

RESULT 1	AK019607	443 bp	musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930441N18:unclassified, full insert sequence.
LOCUS	AK019607		
DEFINITION	AK019607.1	GI:12859904	
ACCESSION	AK019607		
VERSION	AK019607.1		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (sites)		
JOURNAL	Carninci, P., and Hayashizaki, Y.		
MEDLINE	High-efficiency full-length cDNA cloning		
PUBMED	Meth. Enzymol. 303, 19-44 (1999)		
AUTHORS	2 (sites)		
JOURNAL	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
MEDLINE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
PUBMED	Genome Res. 10 (10), 1617-1630 (2000)		
AUTHORS	3 (sites)		
JOURNAL	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hatada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
MEDLINE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)		
AUTHORS	20530913		

PUBMED 11076861
 4 (sites)
 REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5 (bases 1 to 443)
 AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C.,
 Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
 Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
 Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
 Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
 Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
 Schiraldi, L., Shibata, K., Shibata, Y., Shinagawa, A., Shinkai, T.,
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
 Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamana, I.,
 Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
 Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
 URL: http://genome.gssc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT Please visit our web site (http://genome.gssc.riken.go.jp/) for
 further details.
 FEATURES
 source
 1. 443
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="MGI:1912078"
 /db_xref="taxon:10090"
 /clone="1930441N18"
 /sex="male"
 /issue_type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 1. 443
 /note="evidence:NAS
 unclassifiable"
 misc_feature
 143 a 103 c 98 g 99 t
 BASE COUNT
 ORIGIN
 Query Match 1.18; Score 21; DB 11; Length 443;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 299 gaagactatctcaagcaaacg 319
 |||||||||||||||||||||
 Db 131 GAAGACTATCTCAAGCAAGC 151

RESULT 2
 AM098474/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 SOURCE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 458 bp mRNA linear EST 22-OCT-1999
 ga03906.x1 Moss EST library CPU Ceratodon purpureus cDNA clone
 PEP_SOURCE_ID:CPU011211 3', mRNA sequence.
 AM098474
 AM098474.1 GI:6068785
 EST.
 Ceratodon purpureus.
 Ceratodon purpureus
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Bryophyta:
 Bryopsida: Dicranidae: Dicranales: Ditrichaceae: Ceratodon.
 1 (bases 1 to 458)
 Quatrano, R., Bashlades, S., Cove, D., Cumling, A., Knight, C., Clifton
 S., Marra, M., Hillier, L., Page, D., Martin, J., Wylie, T., Underwood
 K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,
 Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
 Waterston, R. and Wilson, R.
 Leeds/Mash v Moss EST Project
 Unpublished (1999)
 Other-ESTs: ga03906.y1
 Contact: Ralph Quatrano
 Leeds/Mash v Moss EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Libraries were constructed by Dr. Stavros Bashlades as part of the
 Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
 Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
 University Genome Sequencing Center for information on obtaining a
 clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
 Seq primer: -400P from Gibco
 High quality sequence stop: 449.
 FEATURES
 source
 1. 458
 /organism="Ceratodon purpureus"
 /db_xref="taxon:3225"
 /clone="PEP_SOURCE_ID:CPU011211"
 /clone_lib="Moss EST library CPU"
 /lab_host="DH10B"
 /issue_type="protonemata: 7 day old tissue"
 /note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2:
 XhoI; Construction of the cDNA library was carried out
 using Stratagene's 'UniZAP - cDNA synthesis kit'. cDNA was
 constructed using an oligo dt primer/linker that contains
 a XhoI site within it. Following ds cDNA synthesis, EcoRI
 adapters were ligated to the blunt ends and sample was
 digested with XhoI. The result is cDNA with an EcoRI
 sticky end on one side and a XhoI sticky end on the other.
 This cDNA was ligated directionally in UniZAP arms. The
 vector is designed containing the bluescript sequence as
 well as lambda DNA and cDNA is cloned within this
 bluescript sequence. The vector was then packaged using
 Gold gigaPackaging extracts. Library was grown in XLBlue
 MRF cells and amplified. The library was excised by mass
 excision using Stratagene's 'Mass excision kit' that uses
 exsist as a helper phage that releases the bluescript
 sequence and circularises it as single stranded plasmids
 that are then packaged (by helper phage) and secreted out
 of the host cell as phagemids. SOLR cells were transformed
 with phagemids and the library was plated out on LB-amp
 plates to select for transformants. Approximately 1,000
 ,000 colonies were grown and recovered. The double
 stranded plasmid library was recovered by using Qiagen
 Midi prep kit. 2 micro grams of each library were used to
 transform DH10B cells by electroporation."
 141 a 51 c 121 g 144 t 1 others
 BASE COUNT
 ORIGIN
 Query Match 1.18; Score 21; DB 9; Length 458;

Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 583 tcatctacattcttcttica 603
|||||
Db 350 TCATCTCATCTTCTTCTTCA 330

RESULT 3

LOCUS AM680975 534 bp mRNA linear EST 19-JUL-2000
DEFINITION WS1_9_H08.bl_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA, mRNA
sequence.

ACCESSION AM680975
VERSION AM680975.1 GI:7554901
KEYWORDS EST.

SOURCE sorghum.
ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.

TITLE An EST database from Sorghum: water-stressed plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmprratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: JEN REV
High quality sequence stop: 531
POLYA=NO.

FEATURES

Source

1..534
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: Lambda Zap; Site_1: XhoI;
Site_2: EcoRI; The library was made from poly-A RNA in the
cloning vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 168 a 143 c 108 g 115 t
ORIGIN

Query Match 1.1%; Score 21; DB 9; Length 534;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 509 gcatctccctatctctgtcatc 529
|||||
Db 84 GCATCTCCCTATCTCTGTCA 104

RESULT 4

LOCUS AM564895 582 bp mRNA linear EST 19-JUL-2000
DEFINITION LG1_312_D09.bl_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
sequence.

ACCESSION AM564895
VERSION AM564895.1 GI:7218773
KEYWORDS EST.

SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 582)

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
TITLE An EST database from Sorghum: light-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmprratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: JEN REV
High quality sequence stop: 566
POLYA=NO.

FEATURES

Source

1..582
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)
seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT 174 a 159 c 135 g 114 t
ORIGIN

Query Match 1.1%; Score 21; DB 9; Length 582;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 509 gcatctccctatctctgtcatc 529
|||||
Db 183 GCATCTCCCTATCTCTGTCA 203

RESULT 5

LOCUS A2538423 909 bp DNA linear GSS 14-NOV-2000
DEFINITION ENTFT26TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.

ACCESSION A2538423
VERSION A2538423.1 GI:11143345
KEYWORDS GSS.

SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.

REFERENCE Eukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 909)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library

JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@fslsrlg.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun

High quality sequence start: 18
High quality sequence stop: 770.

FEATURES

Source

1..909
Location/Qualifiers

```

/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1. Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described from Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      288 a      111 c      170 g      340 t
ORIGIN

Query Match      1.1%; Score 21; DB 12; Length 909;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 556 taattgaagacactaagt 576
|||||
Db 875 TAATTGAAGAACACTAAGT 895

RESULT 6
LOCUS      BG959111      216 bp      mRNA      linear      EST 12-JUN-2001
DEFINITION PM4-CT0806-180301-003-d02 CT0806 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG959111
VERSION     BG959111.1 GI:14377282
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 216)
AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE     20202663
COMMENT     Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&ct=2-PM4-CT0806-
180301-003-d02&ts=2001-03-18&tt=1)
Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 216.
Location/Qualifiers
1. .216
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0806"
/dev_stage="Adult"

FEATURES
SOURCE

```

```

/note="Organ: colon; Vector: puc18; site_1: SmaI; site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      62 a      44 c      58 g      52 t
ORIGIN

Query Match      1.0%; Score 20; DB 10; Length 216;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 ttgaaagagtcagatg 30
|||||
Db 168 TTTGAAAAGTCATGAT 187

RESULT 7
LOCUS      AA669611/c      224 bp      mRNA      linear      EST 20-NOV-1997
DEFINITION IMAGE:856928 3', mRNA sequence.
ACCESSION  AA669611
VERSION     AA669611.1 GI:2631110
KEYWORDS   EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 224)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theisling,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-NCI human EST Project
JOURNAL     Unpublished (1997)
COMMENT     Contact: Wilson Rk
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. EP from Amersham.
Location/Qualifiers
1. .224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:856928"
/clone_lib="Stratagene ovary (#937217)"
/sex="Female"
/dev_stage="49 year old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT.
Total ovary tissue, normal, caucasian. Average insert
size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGACGAG 3' -3' adaptor sequence: 5'
CTCGACGTTTCTTTTCTTTTCTTTT 3'"
BASE COUNT      54 a      47 c      25 g      98 t
ORIGIN

Query Match      1.0%; Score 20; DB 9; Length 224;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 catgaatacttctagag 83
|||||

```

Db 176 CATTGAATCTGCTAGAG 157

RESULT 8 326 bp mRNA linear EST 06-SEP-2000
LOCUS BE649446

DEFINITION UI-M-BH2.2-son-b-10-0-UI.r1 NIH_BMAP_M.S3.2 Mus musculus cDNA clone

ACCESSION BE649446

VERSION BE649446.1 GI:9975235

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Bonaldo,M.F., Lennon,G. and Soares,M.B.
1 (bases 1 to 326)

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements

Seq primer: M13 Reverse.

FEATURES
source Location/Qualifiers
1..326
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_uid="UI-M-BH2.2-son-b-10-0-UI"
/clone_id="NIH_BMAP_M.S3.2"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTZ19-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M.S3.2 library is a subtracted library of a
series, ultimately derived from a mixture of individually
tagged normalized libraries from ten regions of the mouse
brain (cerebellum, brain stems, olfactory bulbs,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus) after a series of
subtractions to reduce the representation of cDNAs from
which ESTs had already been generated. The following
serially subtracted libraries were generated in this
process: NIH_BMAP_M.S3.2, NIH_BMAP_M.S2, NIH_BMAP_M.S1.
The subtracted library (NIH_BMAP_M.S3.2) was constructed
as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2
clones from which 3' ESTs had been derived was used as a
driver in a hybridization with the NIH_BMAP_M.S2 library
in the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the
NIH_BMAP_M.S3.2 library. This procedure has been
previously described (Bonaldo, Lennon and Soares, Genome
Research 6:791-806, 1996)"

BASE COUNT 99 a 72 c 82 g 73 t

ORIGIN

Query Match 1.0%; Score 20; DB 10; Length 326;

Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1423 ctccatcaggaagccttag 1442
|||||

Db 77 CTCACATCAGGAGCCTTAG 96

RESULT 9 370 bp mRNA linear EST 08-NOV-2001
LOCUS AV855875/c

DEFINITION AV855875 Nori Satoh unpublished cDNA library, egg cDNA

ACCESSION AV855875

VERSION AV855875.1 GI:16843399

KEYWORDS EST.

SOURCE Ciona intestinalis.

ORGANISM Ciona intestinalis.
Ciona intestinalis.
Eukaryota; Metazoa; Chordata; Olfrochordata; Ascidiacea; Enterogona;
Phlebobranchia; Clonidae; Ciona.

REFERENCE 1 (bases 1 to 370)
Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
Expressed genes in Ciona intestinalis
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

TITLE Location/Qualifiers

JOURNAL 1..370
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone_id="rcieg22b12"
/clone_uid="Nori Satoh unpublished cDNA library, egg"
/tissue_type="whole animal"
/dev_stage="egg"

BASE COUNT 127 a 48 c 62 g 130 t 3 others

ORIGIN

Query Match 1.0%; Score 20; DB 9; Length 370;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 124 ctggagcttattgtgta 143
|||||

Db 57 CTTGGACTTTATGTGCTA 38

RESULT 10 393 bp DNA linear GSS 22-SEP-1998
LOCUS AQ124704/c

DEFINITION HS-2266.AL.D10.MF CIT Approved Human Genomic Sperm Library D Homo

ACCESSION AQ124704

VERSION AQ124704.1 GI:3501870

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 393)
Mahalnis,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahalnis GG, Wallace JC, Hood L

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Sequence Tagged Connector

Plate: 2266 row: G column: 19

Class: BAC ends

High quality sequence stop: 393.

FEATURES

Location/Qualifiers

1..393

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_1id="2266 Col-19 Row-G"

/clone_1lb="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC1; BAC Clones in E-Coli DH10B"

BASE COUNT

132 a 64 c 69 g 128 t

ORIGIN

Query Match 1.0%; Score 20; DB 12; Length 393;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 gtacgcaatctacacagcagc 226

Db 356 GTACGCATCTACACAGCAG 337

RESULT 11

A1159433

LOCUS

A1159433 433 bp mRNA linear EST 02-OCT-1998
v279b05.r1 Soares_mammary_gland_NbMNG Mus musculus cDNA clone
IMAGE:1332657 5' similar to TR:062146 062146 SMALL GTP-BINDING
PROTEIN ASSOCIATED PROTEIN ; mRNA sequence.

ACCESSION

A1159433

VERSION

A1159433.1 GI:3692615

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 433)
Marrin, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Weising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

AUTHORS

The WashU-HMNI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNC; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:692201

TITLE

Unpublished (1996)

JOURNAL

Unpublished (1996)

COMMENT

Unpublished (1996)

FEATURES

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

/sex="male"

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 415.

Location/Qualifiers

1..433

source

/organism="Mus musculus"

Strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1id="1332657"

/clone_1lb="Soares_mammary_gland_NbMNG"

</

ORIGIN

Query Match 1.0%; Score 20; DB 10; Length 452;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 177 agctctagtcataagatc 196
|||||
245 AGTCTGTAGTCTATAGATC 226

RESULT 13

LOCUS B1073983 470 bp mRNA linear EST 11-DEC-2001
DEFINITION B1073983 NIBB Mochii normalized Xenopus tailbud library Xenopus
laevis cDNA clone XL110103 5', mRNA sequence.

ACCESSION B1073983
VERSION B1073983.1 GI:17504172
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 470)
AUTHORS Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-I, T. and Kohara Y.

TITLE X.
JOURNAL Expressed genes in X. laevis embryo
COMMENT Unpublished (2001)
Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source 1..470
Location/Qualifiers
1..470
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XL110103"
/clone_lib="NIBB Mochii normalized Xenopus tailbud library"
/tissue_type="whole embryo"
/dev_stage="stage 25"

BASE COUNT 168 a 75 c 122 g 104 t 1 others
ORIGIN

Query Match 1.0%; Score 20; DB 10; Length 470;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 gcttggggaagaagaaga 948
|||||
Db 435 GCTTGGGGAAGAAGAGA 454

RESULT 14

LOCUS B1729956 491 bp mRNA linear EST 20-SEP-2001
DEFINITION 603350062F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5357703 5',
mRNA sequence.

ACCESSION B1729956
VERSION B1729956.1 GI:15706982
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 491)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1909 row: k column: 16
High quality sequence stop: 488.

FEATURES
source 1..491
Location/Qualifiers
1..491
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5357703"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-Sport6; Site_1: Notf.
Site_2: Salt; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

BASE COUNT 140 a 104 c 144 g 103 t
ORIGIN

Query Match 1.0%; Score 20; DB 10; Length 491;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1423 ctacatcaggaagccttag 1442
|||||
Db 289 CTCACATCAGGAAGCCTTAG 308

RESULT 15
LOCUS BM409927/c 496 bp mRNA linear EST 22-JAN-2002
DEFINITION EST584254 tomato breaker fruit Lycopersicon esculentum cDNA clone
CIEG50D8 5' end, mRNA sequence.
ACCESSION BM409927
VERSION BM409927.1 GI:18261557
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE 1 (bases 1 to 496)
Alcala, J., Vrebalov, J., White, R., Vision, T., Karamycheva, S.A., Tsai
J., Bougri, O., Kirkness, E., Uterback, T., Van Aken, S., Rinning
C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.

FEATURES
source 1..496
Location/Qualifiers
1..496
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CIEG50D8"

```
/clone_lib="tomato breaker fruit"  
/tissue_type="pericarp"  
/dev_stage="breaker"  
/lab_host="SOLR"  
/note="Vector: pBluescriptSKmCuaapt; Site_1: EcoRI;  
Site_2: XhoI; supplier: Boyce Thompson Institute;  
sequencing: The Institute for Genomic Research. Fruit  
were harvested at the breaker stage (first sign of  
lycopene accumulation on the blossom end of fruit). Fruit  
were cut in half and the seeds and locules were discarded  
prior to freezing the pericarp."  
BASE COUNT      149 a      88 c      127 g      132 t  
ORIGIN
```

```
Query Match      1.0%; Score 20; DB 10; Length 496;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 177 agctctgtagtcataagatc 196  
|||||  
Db 245 AGTCTGTAGTGCATATAGATC 226
```

Search completed: August 20, 2002, 02:30:56
Job time: 4901 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 01:14:20 ; Search time 55.18 Seconds
(without alignments)
8489.001 Million cell updates/sec

Title: US-09-662-812-1
Perfect score: 1907
Sequence: 1 gfgcgtgattgaaag.....ctcacaagagcgtgtaa 1907

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/pdata/2/1na/5A.COMB.seq:*
2: /cgn2_6/pdata/2/1na/5B.COMB.seq:*
3: /cgn2_6/pdata/2/1na/6A.COMB.seq:*
4: /cgn2_6/pdata/2/1na/6B.COMB.seq:*
5: /cgn2_6/pdata/2/1na/PCTUS.COMB.seq:*
6: /cgn2_6/pdata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	0.9	401	US-09-221-298-45	Sequence 45, Appl
2	18	0.9	1030	US-08-858-207A-102	Sequence 102, App
3	18	0.9	10091	US-09-058-489-34	Sequence 34, Appl
4	17	0.9	454	US-08-680-326-118	Sequence 118, App
5	17	0.9	957	US-08-858-207A-104	Sequence 104, App
6	17	0.9	2057	US-09-082-089-4	Sequence 4, Appl
7	17	0.9	2190	US-09-082-089-1	Sequence 1, Appl
8	17	0.9	2625	US-08-759-945-1	Sequence 1, Appl
9	16	0.8	57	US-08-888-381-8	Sequence 8, Appl
10	16	0.8	84	US-08-182-175A-27	Sequence 27, Appl
11	16	0.8	84	US-08-182-175A-28	Sequence 28, Appl
12	16	0.8	84	US-08-474-633A-68	Sequence 68, Appl
13	16	0.8	84	US-08-474-633A-69	Sequence 69, Appl
14	16	0.8	84	US-08-474-633A-69	Sequence 69, Appl
15	16	0.8	84	PCT-US92-06412-27	Sequence 27, Appl
16	16	0.8	84	PCT-US92-06412-28	Sequence 28, Appl
17	16	0.8	175	US-08-182-175A-58	Sequence 58, Appl
18	16	0.8	175	US-08-474-633A-76	Sequence 76, Appl
19	16	0.8	175	PCT-US92-06412-58	Sequence 58, Appl
20	16	0.8	273	US-09-050-159-126	Sequence 126, App
21	16	0.8	399	US-08-894-699-18	Sequence 18, Appl
22	16	0.8	399	US-09-444-410-18	Sequence 18, Appl
23	16	0.8	681	US-08-329-681A-14	Sequence 14, Appl
24	16	0.8	782	US-08-261-825-1	Sequence 1, Appl
25	16	0.8	782	US-08-719-124-1	Sequence 1, Appl
26	16	0.8	782	PCT-US95-07748-1	Sequence 1, Appl
27	16	0.8	930	PCT-US95-07748A-1	Sequence 1, Appl
27	16	0.8	930	US-08-936-165A-74	Sequence 74, Appl

C	28	16	0.8	1028	1	US-08-029-328-1	Sequence 1, Appl
	29	16	0.8	1400	2	US-08-481-658B-43	Sequence 43, Appl
	30	16	0.8	1400	2	US-08-477-504A-43	Sequence 43, Appl
	31	16	0.8	1400	2	US-08-485-756A-43	Sequence 43, Appl
	32	16	0.8	1400	2	US-08-485-862B-43	Sequence 43, Appl
	33	16	0.8	1400	3	US-08-787-739-43	Sequence 43, Appl
	34	16	0.8	1400	3	US-08-487-077A-43	Sequence 43, Appl
	35	16	0.8	1400	3	US-08-485-863A-43	Sequence 43, Appl
	36	16	0.8	1400	4	US-08-485-049D-43	Sequence 43, Appl
	37	16	0.8	1400	4	US-09-178-115-43	Sequence 43, Appl
	38	16	0.8	1400	4	US-09-177-776-43	Sequence 43, Appl
C	39	16	0.8	1480	1	US-08-484-105-19	Sequence 19, Appl
C	40	16	0.8	1480	1	US-08-484-106-19	Sequence 19, Appl
C	41	16	0.8	1496	4	US-09-384-212-1	Sequence 1, Appl
C	42	16	0.8	1552	3	US-08-948-705-1	Sequence 1, Appl
C	43	16	0.8	1691	3	US-08-948-564-3	Sequence 3, Appl
C	44	16	0.8	1692	4	US-08-858-207A-39	Sequence 39, Appl
C	45	16	0.8	1789	3	US-08-948-705-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-09-221-298-45
: Sequence 45, Application US/09221298
: Patent No. 6284241
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
: FILE REFERENCE: 210121.471
: CURRENT APPLICATION NUMBER: US/09/221,298
: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: Fastseq for Windows Version 3.0
: SEQ ID NO 45
: LENGTH: 401
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: (212)
: OTHER INFORMATION: Where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (224)
: OTHER INFORMATION: Where n is a, c, g or t
US-09-221-298-45

Query Match 0.9%; Score 18; DB 4; Length 401;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 152 aagagatgcaaaaatt 1766
1749 aagagatgcaaaaatt 1766
|||||
152 aagagatgcaaaaatt 169

RESULT 2
US-08-858-207A-102/c
: Sequence 102, Application US/08858207A
: Patent No. 6348328
: GENERAL INFORMATION:
: APPLICANT: Black, Michael
: APPLICANT: Hodgson, John
: APPLICANT: Knowles, David
: APPLICANT: Nicholas, Richard
: APPLICANT: Stodola, Robert
: TITLE OF INVENTION: NO. 6348328el Compounds
: NUMBER OF SEQUENCES: 552
: CORRESPONDENCE ADDRESS:
```

ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 1030 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-102

Query Match 0.9%; Score 18; DB 4; Length 1030;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 aaactctgacgactta 493
|||||
Db 369 AAAACTCTGACGACTTA 352

RESULT 3
US-09-058-489-34
Sequence 34, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WHI97-089A
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 34
LENGTH: 10091
TYPE: DNA
ORGANISM: Human
US-09-058-489-34

Query Match 0.9%; Score 18; DB 3; Length 10091;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1439 ttagatctactttccaa 1456

Db 8676 ttagatctactttccaa 8693
|||||

RESULT 4
US-08-680-326-118
Sequence 118, Application US/08680326
Patent No. 5925733
GENERAL INFORMATION:
APPLICANT: ROSE, TIMOTHY M.
APPLICANT: BOSCH, MARINX
APPLICANT: STRAND, KURT
APPLICANT: TODARO, GEORGE J.
TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
NUMBER OF SEQUENCES: 152
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,326
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Schliff, J. Michael
REGISTRATION NUMBER: 40,253
REFERENCE/DOCKET NUMBER: 29938-20001.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 454 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 2..454
US-08-680-326-118

Query Match 0.9%; Score 17; DB 2; Length 454;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 aaactctgacgactt 492
|||||
Db 139 AAAACTCTGACGACTT 155

RESULT 5
US-08-858-207A-104/c
Sequence 104, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328el Compounds

NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 957 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-104

Query Match 0.9%; Score 17; DB 4; Length 957;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 tgcacagcttgcaga 737
Db 423 TGCACAGCTTGCAGA 407

RESULT 6
US-09-082-089-4/c
Sequence 4, Application US/09082089
Patent No. 6100060
GENERAL INFORMATION:
APPLICANT: BARNES, MICHAEL
APPLICANT: TESTA, TANIA
OPERATING SYSTEM: DOS
TITLE OF INVENTION: No. 6100060e1 Compounds
NUMBER OF SEQUENCES: 5
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,089
FILING DATE: 20-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9710737.9
FILING DATE: 23-MAY-1997
APPLICATION NUMBER: GB 9803981.1
FILING DATE: 25-FEB-1998
APPLICATION NUMBER: GB 9804007.4
FILING DATE: 25-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F.
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-30166
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2057 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-082-089-4

Query Match 0.9%; Score 17; DB 3; Length 2057;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 907 agcagagctcaagctt 923
Db 1924 AGCAGAGCTCAAGCTT 1908

RESULT 7
US-09-082-089-1/c
Sequence 1, Application US/09082089
Patent No. 6100060
GENERAL INFORMATION:
APPLICANT: BARNES, MICHAEL
APPLICANT: TESTA, TANIA
OPERATING SYSTEM: DOS
TITLE OF INVENTION: No. 6100060e1 Compounds
NUMBER OF SEQUENCES: 5
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,089
FILING DATE: 20-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9710737.9
FILING DATE: 23-MAY-1997
APPLICATION NUMBER: GB 9803981.1
FILING DATE: 25-FEB-1998
APPLICATION NUMBER: GB 9804007.4
FILING DATE: 25-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F.
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-30166
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701

TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2190 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-082-089-1

Query Match 0.9%; Score 17; DB 3; Length 2190;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 907 agcagcagctcaagctt 923
|||||
Db 2057 AGCAGCAGCTCAAGCTT 2041

RESULT 8

US-08-759-945-1
Sequence 1, Application US/08759945
Patent No. 5834249

GENERAL INFORMATION:

APPLICANT: KAZUAKI, Furukawa

APPLICANT: SUGIMURA, Keijiro

APPLICANT: OHSUKE, Kazuhiko

TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: 699 Prince Street

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/759,945

FILING DATE: 03-DEC-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/303,191

FILING DATE: 08-SEP-1994

APPLICATION NUMBER: JP 5-257881

FILING DATE: 08-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: Rea, Teresa Stanek

REGISTRATION NUMBER: 30,427

REFERENCE/DOCKET NUMBER: 001560-227

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2625 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2625

NAME/KEY: mat_peptide

LOCATION: 118..2625

US-08-759-945-1

Query Match 0.9%; Score 17; DB 2; Length 2625;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 467 catcttggaacttc 483
|||||
Db 2427 CATCTTGGAACCTTC 2443

RESULT 9

US-08-888-381-8
Sequence 8, Application US/08888381
Patent No. 6110703

GENERAL INFORMATION:

APPLICANT: Egel-Mitani, Michi

APPLICANT: Brandt, Jakob

APPLICANT: Vad, Knud

TITLE OF INVENTION: Method For Production of Polypeptides

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6110703 No. 6110703disk of No. 6110703th America, Inc.

STREET: 405 Lexington Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/888,381

FILING DATE: 04-June-1999

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Gregg, Valeta A.

REGISTRATION NUMBER: 35,127

REFERENCE/DOCKET NUMBER: 4840.200-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

TELEX:

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 57 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-888-381-8

Query Match 0.8%; Score 16; DB 3; Length 57;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1353 gatttataagctttt 1368
|||||
Db 33 GTATTATAAGCTTTT 48

RESULT 10

US-08-182-175A-27/C
Sequence 27, Application US/08182175A
Patent No. 5559223

GENERAL INFORMATION:

APPLICANT: Saverio Carl Falco

APPLICANT: Sharon J. Keeler

APPLICANT: Janet A. Rice

TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containin

NUMBER OF SEQUENCES: 113

CORRESPONDENCE ADDRESS:

```
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..84
; OTHER INFORMATION: /product="synthetic oligonucleotide"
; OTHER INFORMATION: /standard_name="SM 98"
;
US-08-182-175A-27

Query Match      0.8%; Score 16; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 tagccttgagttctc 1200
   |||||||
DB 44 TAGCCTTGAGTTCTC 29

RESULT 11
US-08-182-175A-28
; Sequence 28, Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..84
; OTHER INFORMATION: /product="synthetic oligonucleotide"
; OTHER INFORMATION: /standard_name="SM 99"
;
US-08-182-175A-28

Query Match      0.8%; Score 16; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 tagccttgagttctc 1200
   |||||||
DB 44 TAGCCTTGAGTTCTC 59

RESULT 12
US-08-474-633A-68/C
; Sequence 68, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; APPLICANT: COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESSEE: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..84
OTHER INFORMATION: /product= "synthetic"
OTHER INFORMATION: /oligonucleotide= "SM"
OTHER INFORMATION: /standard_name= "SM"
US-08-474-633A-68

Query Match 0.8%; Score 16; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1185 tagcctgagttctc 1200
|||||
44 TAGCCTGAGTTCTC 29

RESULT 13
US-08-474-633A-69
Sequence 69 Application US/08474633A
Patent No. 5773691
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND
APPLICANT: COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
TITLE OF INVENTION: METHODS FOR INCREASING
TITLE OF INVENTION: INCREASING THE LYSINE
TITLE OF INVENTION: AND THREONINE CONTENT
TITLE OF INVENTION: OF THE SEEDS OF PLANTS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
ADDRESS: AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,633A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGEL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..84
OTHER INFORMATION: /product= "synthetic"
OTHER INFORMATION: /oligonucleotide= "SM"
OTHER INFORMATION: /standard_name= "SM"
US-08-474-633A-69

Query Match 0.8%; Score 16; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1185 tagcctgagttctc 1200
|||||
44 TAGCCTGAGTTCTC 59

RESULT 14
PCT-US92-06412-27/C
Sequence 27 Application PC/TUS9206412
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Contain
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06412
FILING DATE: 19920807
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..84
OTHER INFORMATION: /product= "synthetic oligonucleotide"
OTHER INFORMATION: /standard_name= "SM 98"
PCT-US92-06412-27
Query Match 0.8%; Score 16; DB 5; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 tagccttgagttctc 1200
 |||
 Db 44 TAGCCTTGAGTTCTC 29

RESULT 15
 PCT-US92-06412-28
 ; Sequence 28, Application PC/TUS9206412
 ; GENERAL INFORMATION:
 ; APPLICANT: Saverio Carl Falco
 ; APPLICANT: Sharon J. Keeler
 ; APPLICANT: Janet A. Rice
 ; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
 ; NUMBER OF SEQUENCES: 113
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: E.I. du Pont de Nemours and Company
 ; STREET: 1007 Market Street
 ; CITY: Wilmington
 ; STATE: Delaware
 ; COUNTRY: USA
 ; ZIP: 19898
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy Disk
 ; COMPUTER: Macintosh
 ; OPERATING SYSTEM: Macintosh System, 6.0
 ; SOFTWARE: Microsoft Word, 4.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US92/06412
 ; FILING DATE: 19920807
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/743,006
 ; FILING DATE: 9 August 1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Linda Axamethy Floyd
 ; REGISTRATION NUMBER: 33,692
 ; REFERENCE/DOCKET NUMBER: BB-1031
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (302) 992-4929
 ; TELEFAX: (302) 892-7949
 ; TELEX: 835420
 ; INFORMATION FOR SEQ ID NO: 28:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 84 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..84
 ; OTHER INFORMATION: /product="synthetic oligonucleotide"
 ; OTHER INFORMATION: /standard_name="SM 99"
 PCT-US92-06412-28

Query Match 0.88; Score 16; DB 5; Length 84;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 tagccttgagttctc 1200
 |||
 Db 44 TAGCCTTGAGTTCTC 59

Search completed: August 20, 2002, 03:12:04
 Job time: 7064 sec

THIS PAGE BLANK (USPTO)

/translaction="MMSNPRLMKRCFLFASPVLMSGSSADALTTHOEAVKKKNSTYS
HFKSVSGITVIEDGVNLHNHNRIOANKVVENPVMSGSLVLAHGNMNVNRAATVLC
DILEYVEDTSDCLTNRPAMTPEWLEGSMITLTPETIVIRKGIISTEGKNDLUS
GDLEYSLSLSTIGKTLNVCRIPIFLPPEFSIPMEIRKPIPNFGGJGFLGSL
GMSYSPISSRHFSSTFELDFKRGVGMGNLHCSQOVEMENPMKSYAHLRAIIM
AEADRYRLHGDCEFTKHNHSEYHLSDSMEVADIPFNPMFLKMGPRVCCWVN
DNVEGYLTSSVKNFSONOELPYLTLOYPISYINTVLENIYECGYNANFSD
HIGENESSRLIARAPKRLHPICTLSSTGSSLIYSDVEISSRSOLSALQ
LDYRLLHKSXYIQRHIIPEPVFTITRPLAKMEDHITISIDAFHSLNLRKIDT
SVYSKTRPFRPHAKLMTHTHLSNTEKSPKPEKACELPFGKMGVSLDAEMIK
KHCDHNMNIWEMNDGNVAMTLESILHRSYSLICDRENFITLHSPRIDOLDSPLS
DHRNLILGKLFVRPHPCWNYRLSLRYGMHRQDPNRYLEYOMILDTKIFEHMOQLYGYE
RRADSSEFFFLKIDKPKDPF"

BASE COUNT 2399 a 1827 c 1667 g 2375 t
ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 8268;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcttgatttgaataaagtcacatgagatggtttataatgctcaaggtctccatcc 60
Db 4207 GTGGCTTGATTGAAAAAGGTCATGATGTGTTATTAATGTTCAAGGCTCCCTATCC 4266
QY 61 aaacattgaataactgctagagagattgaacatcgatctatggaactatccatcaac 120
Db 4267 AAACATTGAATAACTTCTCTAGAGAGAGTTGAACATCGATCTATGGACTATATCCATCAAC 4326
QY 121 tctctttggaacttatttggtagtctccatctctcttcttggcttaaatccctgaagtc 180
Db 4327 TCTCTTTGGACTTTATTGTGTACTCTCCATTTCTTCTTCTTAATTCCTTGAGTC 4386
QY 181 tgttagtcataaagtccttataataagtaagcaatctacaagaagccttagacaata 240
Db 4387 TGTAGGTCATTAAGATCTTTATATTAATAGTACCAATCTACACAGCAGGCTTAGCAACATA 4446
QY 241 tctgaagcctcctagatgctcctacagtgatcactcttctgttttaagaanaaatccgaga 300
Db 4447 TCTGGAAGCTCTATAGTCCATAGGCTGATCATGACTTCTTCTTTTAAGAAAAATCCGAGA 4506
QY 301 agactatcctaagaacacatccatccatcgaatccgaaactagaanaaacacatcat 360
Db 4507 AGACTATCTCAAGCAAGCAATCCACTCCTCAGATCCGCAAACTGAAAAAGCACATCAT 4566
QY 361 tggagcagcctgagcagatctcttaagaagccttgagcgtgtctccccaagctatgagaac 420
Db 4567 TGGAGCAGGCTGGGGGATCTTCAGAAAGCTTGAGCGTGTCTCCCAAGGTATGAGAAC 4626
QY 421 tgcagacccctgagcagcactactggtttatcgcgcagctcgaagacatcttgggaanaac 480
Db 4627 TGCAGACCCCTGAGAGAGCTACTGGTTTATCGGAGTCTCAGGACATCTTGGGAANAAC 4686
QY 481 tcttgagcactactggttttaagcctttagatctccatccttgcatactccgcttagaagc 540
Db 4687 TCTTGAGCAGACTTACTGTTTAAAGCTTTAGCATCTCCCTATCTGTCATCCGCTTAGAGAC 4746
QY 541 ccgctatagacttgctaatcttgaagaacactaaagtaattgatactatcatcttcat 600
Db 4747 CGGCTATAGACTTGTAAATTTGAAGAACACTTAAAGTATGATCATCTACATCTTTCAT 4806
QY 601 tctaagcctccgaaanaatacgaatgctatctgagcagatcttctaagcttggagac 660
Db 4807 TCTAAGCTTCCCGAAGAAATCCAAATCCATCTGCGGCAATATTCCTAGCGTTGGAGAC 4866
QY 661 tgaagaatctgactatatactcggagctcttagctgcgaagaanaagcgcattcgag 720
Db 4867 TGAAGAATCTGATCTTATTTGGGATCTCTTAGCTGCGCAAGAAAGCCGATTCGAG 4926
QY 721 tgcacagacttgacagatcggaataacacaanaagccttcttccgacacttagaag 780
Db 4927 TGCACAGACTTTCAGATCGAGATACCAACAAAAAGCTTCTTCCGACACTTAGA 4986

QY 781 ttgttaacgagtgctgctctcctcaagaatcaagaactatctttagctttagggaagct 840
Db 4987 TTTGCTTAACGAGTGGCTGCTCCCAAGATCAACAACCTATTCTTTATGCTTTAGGAAGCT 5046
QY 841 taaggaatgctcagagctactaataataaanaagaattgcaagccttgatgtagt 900
Db 5047 TAAGATGCTCAGACTCTACAAATATANAAGCAATGCGAAGCCTGATGTGATGT 5106
QY 901 caacttaagcagcagcttcaagctttaatgctttggggaagaagaagcgcctctccgc 960
Db 5107 CACTTACAGACAGCTCAAGCTTTAATGCTTTGGGGAAGAAGAGAGAGCTCTCCCGT 5166
QY 961 gataaaaaagaagcacttgagagagcgcctccgaccccttgatgctcctaagcctacc 1020
Db 5167 GATAAAAAAGCAGCCTTGAGAGAGGAGGCGCTCGAGCCCTGATAGCTTAGCGCATCTAC 5226
QY 1021 ctctgagaatgagatctccgacttgccctccgatatctccttaaaaaactaagaacaggaagc 1080
Db 5227 CTCTGAGATGAGGATTCGGATTCGCCCTCCGATATTCCTAAAACTAAGAACAGGAAAGC 5286
QY 1081 caagttgaatgtagcttagctctcttagagttgaagtgtagacacccctaaactatgga 1140
Db 5287 CAAGTTGAATGTAGCTTAGCTCTCTTAGAGTTAGGAGGTGACACCCCTAAACTACTGGA 5346
QY 1141 ataatatcaggaagccttgctccaaaccatataatgaagccttagccttagagttctc 1200
Db 5347 ATAACTTACCGAAAGGCTTGTCACACCATTAATATAGACTTAGCGCTTAGAGTTCTC 5406
QY 1201 taagggcgctacttacaanaattgaaagcgggtgaacatactagtcctcctaagaatcccca 1260
Db 5407 TAAGGGCGTACTTTACAAATTTGGAACGGGTGACACTATAGTCCCTCAAGATCCCA 5466
QY 1261 ggagaagggaaggttgctctccacaaccggaagctcttgaagaagaatccttaagttct 1320
Db 5467 GGAGAGGGAAGGTTGCTCTCCACAAACCGAGGCTTGAAGAAGAGATCTTACGTTCT 5526
QY 1321 ctccgcctactcctaagaagcttccctccctgattatagaagcttttggagatcgaga 1380
Db 5527 CTTCGCTTACTTAAGAAGGTTACTCCCTGTATTATATAGCTTTTGGAGTCAACA 5586
QY 1381 aactcagcttgccactactgagatctctctttaaagtaacacccctcaactcagaagcct 1440
Db 5587 AACTCAGCTTGCCACTGCTGAGATTTCTTTTAAAGTACACCCACATCAGAAAGCCTT 5646
QY 1441 agatctactcttccaaagctcgaaagctccttggaacactatcaatccgcgctatgaga 1500
Db 5647 AGATCTACTTTTCCAAAGCTGCGAAGCTTCCGGAAGACTTATCATCCGCCCTATGCAGA 5706
QY 1501 tctgtctattatataatcccaagaagctcctgaaanaaaagcttctccatgatatgc 1560
Db 5707 TCTTGCTATTATATATCTCACCAAGATCTGAAAAAAGGTTCTCTCCATGATTTATGC 5766
QY 1561 aaaaaagctaatcaagaanaacctgctcatctgttgagacaggaanaacaaagcccatcc 1620
Db 5767 AAAAAAGCTAATTCAGAAACCTTGTTATTTGTGACACGGAACCAAGAACCCCATCC 5826
QY 1621 caagcagccctatctacgttatcagtgtaacgcccagaagccgtacgaagctcatgttga 1680
Db 5827 CAGCATGCCCTATCTAGTATAGTATCAGTACACCCCGAAGCCGTCAGAAAGCTATGTTGA 5886
QY 1681 tattctagaagaacatgaacacactcgaagcttccggaagatatccgttatgtatacaact 1740
Db 5887 TATTCTAGAGACACTTACCCACCTCGAAGTCTTCGGAAGATATCGTTTATTTGATACACT 5946
QY 1741 gatgaaggaaggaatgcaaaaaattcccaagtccttgagagccttactgaataaatgt 1800
Db 5947 GATGACGGAAGGAGATCAAAAAATTTTCCAGTCTTGACAGGCTTACTCAATAAAAATGT 6006
QY 1801 ggaagtaaccccaactcagctctatgaagaagcttgctctcatcttctagctccctgttc 1860
Db 6007 GGAAGTAACCCCAACCTATGCTTATGAAAGCTGCTTCTTATTTCTGACTTCCCTTTGTTC 6066
QY 1861 ttatgggttccctagctgtagctttagctatcaagaagcgtgtagaa 1907

```
|||||
Db      6067 TTATGGTCTCAGTCGATGCTTTGACGCATCAAGAGCGCTGCAAA 6113
RESULT 3
LOCUS   AE001587                                16448 bp  DNA   linear   BCT 01-DEC-2000
DEFINITION Chlamydia pneumoniae section 3 of 103 of the complete genome.
ACCESSION AE001587 AE001363
VERSION   AE001587.1 GI:4376271
KEYWORDS
SOURCE    Chlamydia pneumoniae CWL029.
ORGANISM  Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 16448)
AUTHORS   Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
           Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
           Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL   Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE   99206606
PUBMED    10192388
REFERENCE 2 (bases 1 to 16448)
AUTHORS   Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
           Grimwood,J., Davis,R.W. and Stephens,R.S.
           Direct Submission
JOURNAL   Submitted (01-DEC-1998) Program in Infectious Diseases, University
           of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
Source    1..16448
           /organism="Chlamydia pneumoniae CWL029"
           /strain="CWL029"
           /db_xref="taxon:115713"
misc_feature 172..2957
           /note="Polymorphic Outer Membrane Protein G Family;
           contains frame shift"
gene       172..705
           /gene="pmp_3.1"
           /note="Cp0014"
           615..2957
           /gene="pmp_3.2"
           /note="Cp0015"
           3196..4971
           /gene="pmp_4.1"
           /note="Cp0016"
           3196..5953
           /note="Polymorphic Outer Membrane Protein G Family;
           contains frame shift"
gene       4874..5953
           /gene="pmp_4.2"
           /note="Cp0017"
           6302..7786
           /gene="pmp_5.1"
           /note="Cp0018"
           6302..9139
           /note="Polymorphic Outer Membrane Protein G Family;
           contains frame shift"
gene       7787..9139
           /gene="pmp_5.2"
           /note="Cp0019"
           complement(9380..11467)
           /gene="Cp0020"
           complement(9380..11467)
           /gene="Cp0020"
           /codon_start=-1
           /transl_table=11
           /product="Predicted OMP [leader peptide: outer membrane]"
           /protein_id="A018173.1"
           /db_xref="GI:4376272"
           /translation="MKRCFLFLASPVLMGSSADALPHQEAIVKKNLSYLSHFKSVSGIV
           TIDGVLTINHNRIQAKKYVENVYQSGSLKVAHGVMYVNRKTLVYCYLEYEPT
           DSCLTNGRPAKMPWFLGSGMITLPEPTIVIRKGYISTSGPKKDLCSDDYLETSSD
           SLISGKTYLRCRIPILFLPPFSIMMELPKPPINRGSGTGLSGLSYLQMSAPYSLR
           KHFSSTFFLDSPFKHGVMGFNLHCSQKQYPERVFNMKSYAHRALADMAEADRYFL
|||||
gene      complement(11484..13190)
           /gene="Cp0021"
           complement(11484..13190)
           /gene="Cp0021"
           /codon_start=-1
           /transl_table=11
           /product="Predicted OMP [leader peptide]"
           /protein_id="A018174.1"
           /db_xref="GI:4376273"
           /translation="MGLFHLTLFGLLCSPLISLVAKPPESVGHKILYISTOSTQOAL
           ATYELADAYGDHDFVLRKIGEDYDKOSIHSDPQTRKSTIIGAGLAGSSDALDVIS
           QAMETADPLQOQLIVLSAVSGHIGKTSDDLLFKALSPYVIRLEAAVRLANKNTYVI
           DILHSFTIKLPEETIOCLSAIFLRLETESDAYIRDLAKKSAITSATRLQIGEVQO
           KRLPTLRNLITSASPODEAILYALGKLDGSGSYNIKQOLKPDVDTFLAQAOLI
           ALGEDDALPVIRKQALEERPAIVLRHLRPSIGIPIALPIFKNSAKLVAALA
           LLEAGCDTRKLEYITERLVQPHYNETLAFSGKRTLQMKKVNITIVQDPQERERL
           LSTTRGLEOILFELRPLEKAVLPCTSLASOKTOLATTAISFLSHSBOBALDL
           FOAKRIGPEIIRAVADIATVNLTKDPEKRSIHDYAKKLIQETLLEVTENGRPHPS
           MPYLRQVYPESTKIMLDILETLATSKSESDIRLLIQMTBEDANFPLAOLIKI
           VE"
gene      complement(13172..13762)
           /gene="maf"
           /note="Cp0022"
           complement(13172..13762)
           /gene="maf"
           /codon_start=-1
           /transl_table=11
           /product="Maf-type protein"
           /protein_id="A018175.1"
           /db_xref="GI:4376274"
           /translation="MSLRPLVYSSSPRRKFLKFPVPPVPIPSNFDKSVSYSGPI
           ATYQELAQKAVSELDHSPDCITLITGDTIVSYDGRITFKPKDKADATOMLTTRNO
           THDVVTSIAVLHKGKLTGSETSQISLTMIPDRISYIDTGTLLNCGAYDVCHGGL
           ILKRVGCVYVQGLPIQTLKYLEEINIDLMQYSI"
gene      complement(13791..15383)
           /gene="yjjk"
           /note="Cp0023"
           complement(13791..15383)
           /gene="yjjk"
           /codon_start=-1
           /transl_table=11
           /product="ABC Transporter Protein Arpase"
           /protein_id="A018176.1"
           /db_xref="GI:4376275"
           /translation="MSIVLDKIGKSLGTRILFDVSVENPNCYGLTGPNAGKSTL
           LKILMGAIETPRGSISLPKKVGLIRONIDSEPHDTYLDVYIMONTRLMELALQRDNLY
           LOEFTDALGELGEIEIEIIEENGYSADSAEELLGIGIPNMFPRKMMAMIIDLOF
           RVLLCOLAFGHPALLDLDEPTNLDLISIMLGNPLKDYEGTIVYVSHDRPLNTITIT
           HIADIDYDTIIVPGNVDKVEKMTASREQEDIKSEKKIISOLKEPVAKFGAGSVIA
           SOYQSRIREIKIQPOBELKKSNTQREPIRPLSDSGSKYVLSLEAITKRYGQGVYIH
           PFSLEYIQGDKLIGITGNNGLGKTTLMKLVAGVAPSSGSIKILGHQALICSTFPQNSOV
           LADQGETLEWLRNRYGTINDQIRSVLAKMLFGGDGAKRQIALSGGCTARLLMAG
           MLENNHVLLIDEANNNHLDLESVSALSWAINDYKTAIFVSHDRGLIDCATRLIFD
           KDKITFEFGTMVYDTAGHKQLL"
gene      complement(15438..16376)
           /gene="xerc"
           /note="Cp0024"
           complement(15438..16376)
           /gene="xerc"
           /codon_start=-1
           /transl_table=11
           /product="Integrase/Recombinase"
           /protein_id="A018177.1"
           /db_xref="GI:4376276"
           /translation="MIASIVSFLDYLMKVKASAPHTLRNYCLDNLGKILFLBERGNLA
```

PSPELOATERKXSELPESLFEKHEVMYIAKLJENGKAKRTIKRCLSIKSFAYHC
VIOKILLENPAETIHGPRIPKEPLSPMYADEVLMATPOISKTHGRLDRIMLFPYS
SGRLISELVANNKDEPDLSTHLIRKCKKERILIPYSNAIQIYLNHPDRKRLB
KDPQALFHNFRRISTRSIDRSFQELRNSGSLTTPHTRHATHMLESMDLK
TIOALLHSSLETTTVOYSVKLKQTHOEAHPHA*
BASE COUNT 4710 a 3330 c 3665 g 4743 t
ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 16448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgagcttgatttgaagaaagtcagatggttataagttcaagtcctccatcc 60
Db 13290 GTGCGTTATTTTGAAGAAAGGCCAGATGCTTATTAAGTCAAGCTCCCTATCC 13231
Qy 61 aaacatgaactctgtctagagaggttgaacatcgatctatgtgacattccatcag 120
Db 13230 AAACATTGAAMTACTTGCTAGAGAGTTGAACATGATCTATGAGACTTTCATCTAAC 13171
Qy 121 tctcttggacttatttgtgtagtcctccatctctctgtgtgtaaatccctgaagtc 180
Db 13170 TCTCTTGGACTTTATTTGTAGTCTCCATTTCTCTGTGTATAATTCCCTGAGTTC 13111
Qy 181 ttagatcataagatcccttataataagtaagcaatctacaagagagccttaagaata 240
Db 13110 TGTAGGTATATAGATCTTTTATTAATGATGACCAATCTACAGAGAGGCTTATGCAACATA 13051
Qy 241 tctggaagctctagatgctcagtgatcagactctctgtttaaagaaatctggaga 300
Db 13050 TCTGGAAGCTCTAGATGCTCAGGATCATGACTCTCTGCTTTTAAAGAAAATCGGAGA 12991
Qy 301 agactatctcaagaaagatccatctccatcgacccaatagaaaaagcacatcat 360
Db 12990 AGACTATCTCAAGCAAGAACATCCACTCCTCAGATCCGCAAACTGAAAGACACCATCAT 12931
Qy 361 tggagagagcctggtcgagatcttcaagaagccttggagtgctctcccaagctatggagaac 420
Db 12930 TGGAGAGAGCCTGGGGATCTTCAAGACCTTGGACGTCTCCCAAGCTATGGAAC 12871
Qy 421 tgcgaacccccctgcaagcagctactggtttaaactcgacgtcccaagacatcttggagaac 480
Db 12870 TGCAGACCCCCCTGACAGCAGCTACTGCTTTATCGGCACTCTCAGAGACTCTTGGGAAAAC 12811
Qy 481 tctctgaagcttactgtttaaagctttaagcatctccatctcgttcaatcagccttagaagc 540
Db 12810 TTTCTGAGGACTTACTGTTTAAAGCTTAAAGCTATCCCTATCCTGTCATCCGCTTAGAAGC 12751
Qy 541 cgcctatagacttgtaatttgaagaacacataagctatgtatcaatcttcttcat 600
Db 12750 CGCCTATAGACTTGCTAATTTGAAGAACACTTAAAGTCAATGATCATCTACTTTTCAT 12691
Qy 601 tcatatagcttccgaagaataatccatctcctctcgcgcaatcttccatcgttggagac 660
Db 12690 TCATTAAGCTTCCGAGAAATCAATGCTATCTGCGGCAATATTCCTACGCTTGGAGAC 12631
Qy 661 tgaagaatctgaatgcttatatctggagatctctagcttgcgaagaagaacgagatcggag 720
Db 12630 TGAAGAATCTGATCTTATATTTCCGGAGTCTCTTAGCTGCGCAAGAAAGCGGATTCGGAG 12571
Qy 721 tgcgaagacttgagatctggaagaataccaaacaaagccttcttccgaacttaagaa 780
Db 12570 TGCCACAGCTTTGCAGATCGGAGATACCAACAAAGCGCTTTCTCCGACACTTAAAGAA 12511
Qy 781 ttltgctaacagagtgctctcctcaagaatcaagaagctatcttcaatgttaaggaagct 840
Db 12510 TTTGCTTAAGAGAGTGTGCTCTCTCTCAAGATCAAGAGCTATTCTTTATGCTTTAGGGAAGCT 12451
Qy 841 taagaatggtcagaagctactacaataaaaaagcaattgcgaagaagcctgatgtgatat 900
Db 12450 TAAAGATGGTCTAGAGCTACTACATATAAAAAGCAATTCGAGAGAGCTGATGTGATGT 12391

Qy 901 caatttaagcagagctcaagctttaaattgcttcttgggagaagaagagcgtcttccgt 960
Db 12390 CACTTTACAGAGCTCAAGCTTTTAAATGCTTTGGGGAAGAAAGAGAGCGCTTCCGCT 12331
Qy 961 gataaaaaagcaagacttgaagagcgagccttgaagccttgatctagcagcatcagc 1020
Db 12330 GATAAAAAAGCAAGACTTTGAAGAGCGGCTCGAGCCTGTATGCTTACGGCATCTAC 12271
Qy 1021 ctctgagaatgagatccgatctgcttgcgatatctcctaaanaactaagaacagcgaagc 1080
Db 12270 CTCTGAGATGAGGATTCGAGATTCGCTCCGATTCCTTAAAAAACATAAGAACAGCGAAGC 12211
Qy 1081 caagttgaatgagatcttgaagctctcttgaagttgagtggtgtaaccccttaactactga 1140
Db 12210 CAAGTTGAATGATGCTTTAGCTCTCTTGAAGTTAGGGGTGACACCCCTTAATCTGGA 12151
Qy 1141 atacatccgaagaagcttgcacacacatataataagactcagcctgaagttctc 1200
Db 12150 ATACTTACGGAAGAGCTTGTCCAAACCATTTATATAGACTCTAGCCTTGAAGTTCTC 12091
Qy 1201 taagggcgtaacttacaanaattggaagcgggtgaacatcaatagtcctcaagatcccca 1260
Db 12090 TAAAGGCGGTACTTTACAAATTTGGAACGGGTGAACATCATAGTCCCTCAAGATCCCCA 12031
Qy 1261 ggaagaggaaaggttgcctctcacaacccgaggtcttgaagaagacatccatgaagttct 1320
Db 12030 GGAGAGGGAAGGTTGCTCTCCACAAACCCGAGGTTTGAAGAGCATCTTACGTTTCT 11971
Qy 1321 ctccgcctactcaagaagacttaccctccctgtatataagcttcttggagatcaaga 1380
Db 11970 CTTCGCGCTACTTAAGAAAGCTTACCTCCCTGTATTTATAGCTTTTGGGAGTCAAGAA 11911
Qy 1381 aactcagcttgcacactactgagatcttctttaaagtcacacactcaacatcagaagcctt 1440
Db 11910 AACTCAGCTTCCACTACTACTGCGATTTCTTTTAAAGTCACACCTCAATCAGGAAGCCTT 11851
Qy 1441 agatctacttcttccaagctcggaagctcctcggaagacatcatcgcgctatgaga 1500
Db 11850 AGATCTACTTTTCCAAAGCTGGGAAGCTTCCGGAAGAACCTTATCCCGCTCTATGAGA 11791
Qy 1501 tcttgctattataatctcaccagaagatcctgtaaaaaaacgtctctcctcagttatgac 1560
Db 11790 TCTTGCTATTTATATCTCACCAAGATCTCGAAAAAACGTTCTCTCCATGATATGC 11731
Qy 1561 aaaaaagctaaatcaagaaacactgttatttgttgaacaggaagaaacaaagaccatcc 1620
Db 11730 AAAAAAGCTAATTCAGGAACCTTGTTATTTGTGACACAGGAAAAACCAAGACCCCATGC 11671
Qy 1621 cagcatgctctatctacgttatacagtgatcccccagaagacggtacgaagctcatgttga 1680
Db 11670 CAGCATGCCCTATCTAGCTTATCAAGGTACACCCCGAAGACCGTACGAAGCTCATGTTGGA 11611
Qy 1681 tatctagaagacactagcacaactcgaagctctcgaagatatacgttatatgtataact 1740
Db 11610 TATCTAGAGACATATAGCACTCTGAAGTCTTCGGAAGATATCCGTTATTTGATACAACT 11551
Qy 1741 gatgaaggaagagatgcaaaaaatttccagttccttgcaggttactcatlaaaaaatgt 1800
Db 11550 GATGACGGAAGAGATGCAAAAAATTTCCAGTCTTGCAGGCTTACTCATATAAAAAATGT 11491
Qy 1801 ggaatcccccaactcagctgcttataagaaogttgcttcttattctgcttcccttgc 1860
Db 11490 GGAGTAAACCCCAACCTTACGTTTATGAAGCGTTGCTTCTTATTTCTAGCTTCTTGTGC 11431
Qy 1861 ttatggttcctcagctgagatgcttgcctcaatcaagaagcgtgtgaa 1907
Db 11430 TTATGGTTCTCTCAGCTGATGCTTTGACATCAAGAGCGCTGTGAA 11384

RESULT 4
AX349527 1707 bp DNA linear PAT 06-FEB-2002
LOCUS AX349527
DEFINITION Sequence 50 from Patent WO0202606.

ACCESSION AX349527
VERSION AX349527.1 GI:18615370
KEYWORDS
SOURCE Chlamydomophila pneumoniae.
ORGANISM Chlamydomophila pneumoniae.
REFERENCE
1 (sites)
Bacteria: Chlamydiales; Chlamydiaceae; Chlamydomophila.
Ratl, G. and Grandi, G.
Immunisation against Chlamydia pneumoniae
Patent: WO 0202606-A 50 10-JAN-2002;
Chiron S.P.A. (IT)
FEATURES
source
1..1707
/organism="Chlamydomophila pneumoniae"
/db_xref="taxon:83558"
BASE COUNT 493 a 415 c 340 g 459 t
ORIGIN
Query Match 89.5%; Score 1707; DB 6; Length 1707;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 101 atggagactatccaactcaactcctcttggaacttatttgtagcttccattctct
Db 1 ATGGAGACTATCCACTCACTCTTGACTTTTATGTAGCTTCCCATTTCTCTT
QY 161 gtgtgtaattccctggtctgttagtcataagaaccttataataagaactacaa
Db 61 GTTGCTAAATTCCTGAGCTGTAGTCATAGATCCTTATATATAGTCGAATTCACA
QY 221 cagcagccttaagcaacatctggaagctctagaatgcctacaggtgacatgactctc
Db 121 CAGCAGCCTTAGCAACATATCTGGAAGCTCTAGATGCCCTAGTCAAGATTTCTTC
QY 281 gttttaagaagaatcggagaagactatcacaagaagacatccactccatgaatcga
Db 181 GTTTTAAAGAAAATCGGAGAAAGACTATCTCAAGCAAAAGCATCCACTCGATCCGCAA
QY 341 actagaagaagaccatcatcttggaagcgtgagcaggtcttggaagccttgaagcgtg
Db 241 ACTAGAAAAGACCATCATCTTGAGAGCGCTGCGGAGTCTTAGAAGCCTTGAGACGTG
QY 401 cctcccaagctatggaactgcagacccttcagcagctactggttttaacgcagctc
Db 301 CTCGCCAAGCTATGGAAGTGCAGACCCCTGACAGCAGCTACTGTTTATCGCGAGTC
QY 461 tcaagacactcttggaagaactctgcagactactgtttaagccttgaacatccctat
Db 361 TCAGGACATCTTGGGAAAACCTTGCAGCACTTACTGTTTAAAGCTTATGCACTCCCTAT
QY 521 cctgtatccgttagaagccgcctatagactgtctaatattggaagaacacaaagtatt
Db 421 CCTGTATCCGCTTAGAAGCCGCTATAGACTTCTTAATTTGAAGAACACTAAAGTCTATT
QY 581 gatcatctacatcttcatcatcaagctccggaagaatccaatccaatccatctgcgga
Db 481 GATCATCTACATCTTTCATTCATTAAGCTTCCCGAAGAAATCCAAATGCTTCTGCGGCA
QY 641 atattcctcagcttggaagactggaagactctgaagctatataatccggatctctgaagc
Db 541 ATATTCTACGCTTGGAGACTGAAGAAATCTGATGCTTATATTCGGGATCTCTAGCTGCC
QY 701 aagaagaagcagcttggaagctgcacagcttctgagatcgaagaataccaacaagaagc
Db 601 AAGAAAAGCGGATTCGAGAGCCACAGCCTTGGAGATCGGAGAAATACCAACAAAGAGCG
QY 761 ttcttcogacacttaagaattgttaacagatgctctcctcaagaatcaagaagcatt
Db 661 TTTCTTCGACACTTAGGAATTTGCTAAGAGAGTGGCTCTCTCAAGATCAAGAAAGCTATT
QY 821 cttaagcttaaggaagcttaagaatggtcagaagactactacaataataaagaagaatg
Db 821 |

Db 721 CTTTATGCTTTAGGGAAGCTTAAGGATGTCAGAGCTACTACATATATMAAAGCAATTG
QY 881 cagaagcctatgtagatgtagtcaactttagcagcagctcaagccttaattgcttgggga
Db 781 CAGAAAGCTATGTGATGTACACTTTTAGCGAGCACTCAAGCTTTAATTGCTTTGGGAAA
QY 941 gaagaagacgctctcccgatataaaaaaagcaagcacttgaagagcgagcctcgaagcctg
Db 841 GAAGAGAGACGCTCTTCCCGATTAATAAAGCAAGCACTTGAGAGCGCGGCTCGAGCCCTG
QY 1001 tatgcttaagcgaactacaccttgagatagggaattccgattgcctgcggaattccta
Db 901 TATGCTTACGCACTACCTCTGAGATGGGATTCGATTCGATTCGCGATATTCCTA
QY 1061 aaaaactaagaacagcgaagctgaatgtagcttgaactcctcctagaagtaggtgt
Db 961 AAAACTAGAAACAGCGAAGCCAAAGTTAAATGTACTTTAGCTCTTAGAGTTTAGGCTGT
QY 1121 gaacccctaaactactcgaatacataccgaagaagccttgccaaccacataatagag
Db 1021 GACACCGCTTAAGTACAGTAACATTAACGAAAGGCTTGCCAAACCAATTAATAGAG
QY 1181 acttaagccttgaattctcctaagggcgtaacttacaataatggaagcgggtgaacatc
Db 1081 ACTCTAGCCTTGAGTTTCTCTTAAGGGGCGTACTTACAAAATTGGAAAGCGGTTAACATC
QY 1241 atagtccttaagaatcccccgaaggaaggaaggttgcttcacaaacccgaaggtctgaa
Db 1141 ATAGTCCCTCAAGATCCCGAGAGGGAAGGTTGCTCTCCACAAACCGAGGCTTTGAA
QY 1301 gaagcagatccttaagcttctctccgccttaactcaaaagaagccttccctcgattat
Db 1201 GAGAGATTCCTTAGCTTTCCTTCCGCTACCTAAAGAGTTCACCTCCCTGATTTAT
QY 1361 aagcctttagcgagtcagaagaactcagcttgcacactcgtgattcttltttaagtcac
Db 1261 AAGCTTTGGGAGTTCAGAAAACCAACCTTGCCACTGAGATTTCTTTTAAAGTCAC
QY 1421 accctacataagaagccttagatcacttctcgaagctggaagccttccctggaagcct
Db 1321 ACCTCACATCAGAAAGCTTATAGATCTTTTCCAAAGCTGGAAGCTTCCGGAAGACT
QY 1481 atcatccgcgactatgagaatctgtcatlataatcaccacaagaatcctgaaaaaaa
Db 1381 ATCATCCGGCGCTATGAGATCTGTATTATATCTCACCAAAAGATTCGAAAAAAA
QY 1541 cgttctcctcaatgatatgcaaaaaagcctaattcagaagaacctgttatttgtgacag
Db 1441 CGTTCCTCCATGATTAATGCAAAAAGCTTAATTCAGGAAACCTTGTTATTTGTGACAG
QY 1601 gaaaccaaagaagcccatccagagatgccatctacgttaagatcgaagcgaagaagc
Db 1501 GAAAACCAAAAGACCCATCCAGCATGCTTATCTAGCTTATGAGGTACCCCGAAAAGC
QY 1661 cgtacgaagctcagctttagatatcttagagaagacactgaagccttctcgaagat
Db 1561 CGTAGAAGCGCATGTGGATATTCFAGAGCACTAGCACTCGAAGTCTTCGAGAGAT
QY 1721 atccgcttatatgatacaactgtagcaggaagagatgcataaaatctccagctctgca
Db 1621 ATCCGTTTATGATTAACAAGTATGACGAAAGAGATGCAAAAATTTCCAGTCTTGCA
QY 1781 ggtctactcctaataaattgtgagtaa 1807
Db 1681 GGGTTACTACATAAAATTTGTGAGTAA 1707
RESULT 5
AE001308 11627 bp DNA linear BCF 30-Oct-2000
LOCUS Chlamydia trachomatis section 35 of 87 of the complete genome.
DEFINITION
ACCESSION AE001308 AE001273
VERSION AE001308.1 GI:3328766

D	4512	TTCTGGAAGCTTCAGCAACTTATATCAACAACGATTTTCTCCGATTATTAAGCGCTTAATACGACA	4577
Q	301	agactactcaagaagaagca tccactctccaga tccgcgaactatagaanaagcacatcat	360
D	4572	ATCGTATTTCACAAACGCTTCTCTCTGAGACACCTACATACGTAAAGTGCATATTAT	4633
Q	361	tggagcagcgcttgcggagatctcagaagccttggacgttgcctccaaagctatggaaac	420
D	4632	TGGAGCAGGGCTATCTGGTTCAACAGAGCTTTAGAGTTACTGCTGAGAGCTATAGAAAC	4691
Q	421	tgcagacccccgcgaagctacttglttatcgcgaagctcgaagcatcttgggaaac	480
D	4692	GCAAGATCTGTATGAGCAACACTCTCATTTTTAAATGCTGAAACGACCAATTAAGCAAAA	4755
Q	481	tttcgacgacttactgtttaagctcttagcatctccatccctgtaaccgctatgaaac	540
D	4752	TTCTGACAAACTTTTATTCAAGGATTAAACGCTTCTCATCCGTGATCCGGTTAGAAC	4811
Q	541	cgcctatagacttgcataatttgaagaacacataaagtcataatgatalcatcatcttcat	600
D	4812	TGCTTATCGCTTCCCGTATGAAAAATAGCAAGTAAAGATTAACCTTTATTCCTTTAT	4873
Q	601	tcaatagctccocgaagaatccaagcctatctgtaggaatactccatcgcttggagac	660
D	4872	CTACAAATTCACAAAAATTCAAAACCTAGCGGCAACATATTTCTTACAACTGAAAC	4933
Q	661	tgaagaatctgatactctatatactcggatccttagctgcgaagaanaagcgatcggag	720
D	4932	AGAAAGATCGATGCTTATTATTCATCATTTGCTCTCTTCCCAATTAACCTGACAAGAA	4999
Q	721	tgcacagcttgcagatctggagaataccaacaaaagcgtctcttcgacacttagaa	780
D	4992	CTATGTTGCCATATTATTTAGAGAGTACAAACAAAAGATTTCTTCCAAACCTACGCTC	5053
Q	781	tttgctaaagatgctgctctctcgaatcaagaagctatctttagctttaggaaagt	840
D	5052	TTTACTTACAGTCCCTCTCTTTAGTCAAGAGCGCTTTGTATGCGCTTAGGCAACT	5113
Q	841	taagatagctcagagctactacataaanaaagcaatgcagaagccgagatgtagt	900
D	5112	GGAAAGACTCTGGTAGCTATCTGAAATTAAGCTCTAAGCTCTAAGTCCAAATCTGAAT	5177
Q	901	caacttgcagcagctctcaagctcttaatttgccttgggaaagaagagcgctcttccgt	960
D	5172	AGTACTGCTGACACTCAGACATTATTTCTTAGAGAAAGAAAGAAAGAGCTCTACGAT	5233
Q	961	gataaaaagaagcacttgaaggagcgctcgagccctgatatgctctcagcaatcac	1021
D	5232	CCTTACCACACTTTGGCCACACAAAACCTTCTTCAGCGCTGTATATCCGCAAGTTTCTCTTC	5299
Q	1021	ctcgcgatataggatctccgattgccccgcgcatctccaaanaactaagaacagcgaagc	1080
D	5292	GCAAGAGAGGGTGAGAGGCTTCTTCTTCCAAATCTTTTATACGCCAACACAGAGAAAT	5355
Q	1081	caagtatagtagctctctctctctcttagagtttaggttgcacaccctaaactatgta	1144
D	5352	TAGACTGAATACGTCTTTAGCACTGTGTATCAAGGATGACAGATTCCTCAAGTCTGCA	5411
Q	1141	atacatcaccgaagcgttgcacaccacataataatgagacttagccttgaattctc	1201
D	5412	CTATCTAACGAAATCTTAAAGATTAAGTTCCTCATCGCATTTTCTTACTCATCTACCTTC	5477
Q	1201	taagggcgctacttcaaaaattggaaagcgggtggaacatcatatagctccccaagatccca	1261
D	5472	GACAGGAAAGCTATACAGTTTCTTGGAAGAAATCCACCACTTTCCTCATGAGCCACAGA	5533
Q	1261	ggaagagaaaggttgcctcccaaacccaggtcttgaagagcagatcttaagttct	1321
D	5532	AGACAAATGGAACGTTGGCTATGTATGCGGATGAGGGAAGATACCATCTCTCAGCGTT	5599
Q	1321	cttcgcctactcaagaagcttaccctcccgctlatltaaaagcttttgcgtagtcaaga	1386
D	5592	ACTAAATTCACCAATAGCCCTATCTCTCTTACCTTAGAGCGGATCTCGCTCTCAAAA	5655

OY	1381	aactcaagcttgcacactactcgattctcttlltlaagtcacaccctacatcaggaacgtt	1440
Db	5652	AACTTACTACGAGGATGAAGCTATTGCTTTTATTCGGTACAGCTCATCTCCAGGCACT	5711
OY	1441	agatctactcttccaaagctcgggaagctcttccttgagaacctcattccggccattgcaga	1500
Db	5712	TTCTTTAGTCTCGAANAAGCTGCAATTACACTCGTGGAAACCTTATCATTTCCGGCTTACGCTTA	5771
OY	1501	tcttgcattatataactcaccaaagctcggaaaaaaagcttctctccatgattatgc	1560
Db	5772	TCTACTTTATTATACAAATGACCAAGACCTCTGAGAAAAAAGCTGTGCTTACCGATTATGC	5831
OY	1561	aaaaagcctaactcgggaagccttgatttttgtagcagcggaacaaacgaagcccatcc	1620
Db	5832	TGAAACAATTAAATAGAGGATACCATTTTATTCACAGTGGTGAATAATCGGCTTCCCTTCC	5891
OY	1621	cagcatgcacctacacagcttalcaggtaccaccgaagaacgcgtlaagaagctcatgttga	1680
Db	5892	AAGCTCTTCTTATTATACGGTACCAAGTATGCCCTTGAAACCGCACAAACTTATGCTTACG	5951
OY	1681	tattctagaagaacacgccacctcggaagctcttcggaagatalcgatttattgatacaact	1740
Db	5952	TATTTTGGAAACCTTATGTTCTTCCAAAACGGATAGAGTATACCCCGTTTCTTCTCCCT	6011
OY	1741	gatgcagggaagagatgcgaaaaattccaggtctcgtcagagcttaccataaaattcgt	1800
Db	6012	AATGAAAAAAACCCATTACAAAAAATATCCCGACTTTATCAGGATTTGTTAATAGAAATAGT	6071
OY	1801	ggagtaacc	1809
Db	6072	GGAGTGATC	6080

RESULT	6
AE002331	
LOCUS	AE002331 11765 bp DNA linear BCT 26-MAY-2000
DEFINITION	Chlamydia muridarum, section 61 of 85 of the complete genome.
ACCESSION	AE002331 AE002160
VERSION	AE002331.2 GI:8163269
KEYWORDS	
SOURCE	Chlamydia muridarum.
ORGANISM	Chlamydia muridarum
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
AUTHORS	1 (bases 1 to 11765) Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umeyam,L.A., Uitterback,T., Berry,K., Bass,S., Linher,K., Weidman,U., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
TITLE	Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNAL	
MEDLINE	20150255
PUBMED	10684935
REFERENCE	2 (bases 1 to 11765)
AUTHORS	Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umeyam,L.A., Uitterback,T., Berry,K., Bass,S., Linher,K., Weidman,U., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M. Direct Submission
TITLE	
JOURNAL	
COMMENT	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA On Jun 1, 2000 this sequence version replaced gi:7190662.
FEATURES	
source	Location/Organifiers 1..11765 /organism="Chlamydia muridarum" /db_xref="taxon:83560" /note="synonym: Chlamydia trachomatis Mopn" 93..2552 /gene="TC0623" 93..2552
CDS	

gene
/gene="TC0623"
/note="similar to SP:P36777 PID:41406 PID:429100;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="protease, lon family"
/protein_id="AAF39454.1"
/db_xref="GI:7190663"
/translation="MNSTNTDSQNLDPNASEVEKLDSESAEKEKTDHNPSESLPT
LPLNKKPFPGMAAPDLIAGPHVYLVTLNLAASSKHIGLVTKREDANTLKIGFNOI
HARGVASRLIRLPIEGGSAQVLLSIEIRIRVKNVQYLAQVAAVAKENKKEITFEEI
KASISIVSIIIDPLKLNPLQFKEELQIFLGSHDFEPKELADEFSALTATARELQEEI
LETWDDHIDIDALVLLKKELDLSLOSSINOKIEATTTKQKFEFLKOLKATPELKE
GLEKDDHADLEKEMERLKRDPVQADYVIODEMDKQTLQTESSAEAVARNLDMYI
TVPWGIDYKREYHDLKKAESTLNKRDYKEDIKORIELTISGVLAMGMSITIGC
PTVQGTSTGSRIAKYLHRRKPFRESVCGMRDEAKKGRRTYIGAMRGKLVQALQKQSA
IMPVMLIDEVKIGSSYHGDPSALDLEVDKODLDYLDVRDLNVLTFLTA
NVLDSIPDLDLMEVLRISGVLIEEKIQIAIKYLVPRAREMGI.SAONVSFOEALR
HMINNARAGVARTNENIKYLRKVAIKIYONONKPSKSKRFTIPKNIQDYLRKEP
ITSSDFEYKTEPVGATGLAMTSLGATLYIESYOVSSGKAAMHLTGQADVMKSE
SOLAMVYHLSALERYAPGPRPFKSCVNIHIEGATPGDSAGITWVTSILSLIDVY
PLNINIGMGETLTGRTVIGTIGIKELIARSKNLVLTPEDNRDYDELPATYLKK
GLKHPYTHIVDYKIAEPGV"
2670 . 3035
/gene="TC0624"
2670 . 3035
/gene="TC0624"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39455.1"
/db_xref="GI:7190664"
/translation="MDPLSIIRSFPSPLTSSNIRKSLSEOTERIVRAYLILSLST
LGVISGVSTSLPLGASICLSIALITSCLLVLEPLLPDVARIIYQKRPIPIPKHIA
PSNVLTNRKHSIKLDFKSH"
3306 . 4220
/gene="TC0625"
3306 . 4220
/gene="TC0625"
/note="similar to GB:U00096 SP:047012 PID:1381661
PID:1786603; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Alsa/ElaC family protein"
/protein_id="AAF39456.1"
/db_xref="GI:7190665"
/translation="MSYRGITLIGCSSQOQPTRRHNRHGAYILRWNGEGLRDPGEGTOR
QVYININAPATVYSRIFISHPGDHCTGIGSLMILNDKYSRPHCYHVSQKQKPPRR
LKTSTIYHETIKVLEHPISSBSGIVDFGNFPIEARQDLHDLYDTGKMPITPEDTTKFIS
EKIKAAKGLPIQDELINKGIIKVNNTIIRLEVDASYTRKGDLSIAVNDLSLPCQALIVL
AANARLILCESTYIEEHAHLKNHYHMTAKQAABQAKAEAOQLIITHFSARVNTTBO
FVQAGEIPIPNFAEAEECSYEFPKNN"
4283 . 5230
/gene="TC0626"
4283 . 5230
/gene="TC0626"
/note="Integrase/recombinase, phage integrase family;
identified by match to PFM protein family HMM PF00589"
/codon_start=1
/transl_table=11
/product="Integrase/recombinase, phage integrase family"
/protein_id="AAF73578.1"
/db_xref="GI:8163270"
/translation="MIASFVAFLDYLLKNMKAASPHTLRNTSIDLSLKCFLEKKGEILT
PPPLPSLEDDSSSSQLSFLSTFKENIYLVLEQIQTHSKRTVRRSLAISKAPKC
VYNNQIPEPAMIRGPRLPKPELSPVLTLEUOLVLMASPDLDKVTGFRDCLLEPFS
SGLRISSETLAINRSDIDROSNLIRTCGKGRKEIVPTPKYAVOMLQALYLDHPRAAVE
QDQACFLHNRGKRLSTSIDRKFDQYILKICSGSGTTPIRTITRTIATYHMLERGMDLK
TIQLLGLNLSLETTTITTYHVSMLKKQIHLDEAHPHNLED"
5284 . 6870
/gene="TC0627"

FEATURES	source	Location/Qualifiers
		1..128866
		/organism="Oryza sativa"
		/clivair="Nipponbare"
		/db_xref="taxon:4530"
		/chromosome="2"
		/clone="OJ1212_E12"
BASE COUNT	34028 a 30313 c 29226 g 35194 t	105 others
ORIGIN		
Query Match	2.1%; Score 40.6; DB 2; Length 128866;	
Best Local Similarity	52.7%; Pred. No. 5;	
Matches	88; Conservative 0; Mismatches 79; Indels 0; Gaps 0;	
OY	1332 ccaactacgcgattcttltttaaagtcacacctacacacgaagaagccttagatctt 1451	
Db	5470 CAACCTATTGTTATGACATATTTAAGTCACAGCTTCCTCAGCAGTCGATATCTTTGTG 5411	
OY	1452 tccaaagtcggaagcttcctcggagaacctatcaccgcgcctatgacagatcttctatt 1511	
Db	5410 CTGCTGCTTGGATGCTGATGAACCGACATCTAGAGATCAACAGTCGAATCTTGATGAT 5351	
OY	1512 ataatccaccaagatcctgaaaaaagcttcctcctcatgattat 1558	
Db	5350 TTCTAATTATGCAATTATATTGACATTAAGAATTTCTCTATATGTTATAT 5304	
RESULT 11	AL445687	168108 bp DNA linear PRI 24-NOV-2000
AL445687/c		
LOCUS	Human DNA sequence from clone RP11-567B20 on chromosome 1, complete	
DEFINITION	sequence.	
ACCESSION	AL445687	
VERSION	AL445687.5 GI:11322702	
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
REFERENCE	1 (bases 1 to 168108)	
AUTHORS	Thomas,D.	
TITLE	Direct Submission	
JOURNAL	Submitted (24-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,	
	CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk	
	On Nov 23, 2000 this sequence version replaced gi:11139977.	
COMMENT	During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.	
	This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.	
	The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP. Information on the WORMPEP database can be found at	
	http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at	
	http://www.sanger.ac.uk/HGP/Chr1	
	RP11-567B20 is from the library RPCT-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/	
	VECTOR: pBAC3.6	
	This sequence is the entire insert of clone RP11-567B20.	

FEATURES	
source	Location/Qualifiers
	1. .168108 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /clone_id="RP11-567B20" /clone_1fb="RPCT-11.2" 742. .771
repeat_region	/note="15 copies 2 mer ca 100% conserved" 1032. .1260
misc_feature	/note="MIR repeat: matches 20. .259 of consensus" complement(1277. .1815)
misc_feature	/note="match: STS: Em:G62829" 1546. .2029
repeat_region	/note="match: GSS: Em:AQ454440" 2503. .2702
repeat_region	/note="MIR repeat: matches 34. .238 of consensus" 3036. .3331
misc_feature	/note="AluSg repeat: matches 3. .239 of consensus" complement(4943. .5462)
misc_feature	/note="match: GSS: Em:AQ706543" complement(5017. .5475)
misc_feature	/note="match: GSS: Em:AQ401497" 5481. .6203
repeat_region	/note="match: GSS: Em:AQ377544" 6847. .6981
repeat_region	/note="L2 repeat: matches 2559. .2710 of consensus" 7538. .7661
repeat_region	/note="L2 repeat: matches 2121. .2243 of consensus" 7813. .8115
misc_feature	/note="AluSk repeat: matches 1. .303 of consensus" 8126. .8494
repeat_region	/note="match: STS: Em:G07568" 8341. .8414
repeat_region	/note="THE1B repeat: matches 1. .74 of consensus" 8927. .9222
repeat_region	/note="Aluub repeat: matches 1. .301 of consensus" 9302. .9513
repeat_region	/note="MIR repeat: matches 20. .250 of consensus" 9562. .9760
repeat_region	/note="MIR repeat: matches 48. .249 of consensus" 9849. .9961
misc_feature	/note="L2 repeat: matches 1149. .1258 of consensus" 10016. .10365
repeat_region	/note="match: GSS: Em:AQ276162" 10035. .10160
repeat_region	/note="MT1L repeat: matches 19. .185 of consensus" 10322. .10379
repeat_region	/note="29 copies 2 mer ca 75% conserved" 10500. .10646
repeat_region	/note="L2 repeat: matches 1373. .1544 of consensus" 10731. .10844
repeat_region	/note="MER58 repeat: matches 1. .117 of consensus" 10860. .11134
repeat_region	/note="L2 repeat: matches 1556. .1828 of consensus" 11162. .11490
repeat_region	/note="L2 repeat: matches 1879. .2234 of consensus" 11564. .11790
repeat_region	/note="MSTA repeat: matches 1. .207 of consensus" 11791. .12100
repeat_region	/note="AluSk repeat: matches 1. .312 of consensus" 12101. .12264
repeat_region	/note="MSTA repeat: matches 207. .426 of consensus" 12400. .12519
repeat_region	/note="L2 repeat: matches 2626. .2750 of consensus" 12786. .13164
misc_feature	/note="MSTA repeat: matches 4. .426 of consensus" complement(13355. .13937)
repeat_region	/note="match: GSS: Em:AQ116449" 14332. .14379
misc_feature	/note="MIR repeat: matches 208. .257 of consensus" complement(14897. .15458)
misc_feature	/note="match: GSS: Em:B48614"

```

misc_feature      15471..16013
                   /note="match: GSS: Em:AQ378181"
misc_feature      15475..15980
                   /note="match: GSS: Em:AQ264491"
repeat_region     15558..15643
                   /note="43 copies 2 mer aa 64% conserved"
repeat_region     17157..17216
                   /note="12 repeat: matches 1724..1793 of consensus"
repeat_region     17221..17278
                   /note="29 copies 2 mer ga 69% conserved"
repeat_region     17442..17658
                   /note="MER20 repeat: matches 3..218 of consensus"
repeat_region     17749..18098
                   /note="12 repeat: matches 2353..2690 of consensus"
repeat_region     19071..19210
                   /note="MIR repeat: matches 78..230 of consensus"
repeat_region     19294..19608
                   /note="AluSg repeat: matches 1..313 of consensus"
repeat_region     19912..20041
                   /note="12 repeat: matches 1627..1761 of consensus"
repeat_region     20452..20492
                   /note="MADP1 repeat: matches 1..42 of consensus"
repeat_region     20493..20789
                   /note="12 repeat: matches 1755..2089 of consensus"
repeat_region     21189..21943
                   /note="L1PA15 repeat: matches 4705..5420 of consensus"
repeat_region     21944..22243
                   /note="AluY repeat: matches 1..301 of consensus"
repeat_region     22244..22959
                   /note="L1PA15 repeat: matches 5420..6155 of consensus"
repeat_region     23511..23788
                   /note="AluYb repeat: matches 11..288 of consensus"
repeat_region     23947..24243
                   /note="AluSp repeat: matches 1..298 of consensus"
repeat_region     25369..25670
                   /note="AluX repeat: matches 1..304 of consensus"
repeat_region     26305..26423
                   /note="MIR repeat: matches 20..142 of consensus"
repeat_region     26810..27184
                   /note="12 repeat: matches 1591..2004 of consensus"
repeat_region     27930..28238
                   /note="AluX repeat: matches 2..312 of consensus"
repeat_region     28300..28612
                   /note="AluX repeat: matches 1..311 of consensus"
misc_feature      complement(29577..29949)
                   /note="match: GSS: Em:AQ084247"
repeat_region     31407..33048
                   /note="L1M4 repeat: matches 80..1764 of consensus"
repeat_region     33049..33348
                   /note="AluX repeat: matches 1..298 of consensus"
repeat_region     33349..33598
                   /note="L1M4 repeat: matches 1764..2005 of consensus"
repeat_region     33689..33770
                   /note="Alu repeat: matches 227..308 of consensus"
repeat_region     33908..34591
                   /note="L1M4 repeat: matches 2257..2599 of consensus"
repeat_region     34593..34725
                   /note="AluY repeat: matches 2..134 of consensus"
repeat_region     34759..35417
                   /note="L1M4 repeat: matches 2614..3307 of consensus"
repeat_region     35415..36185
                   /note="L1PA13 repeat: matches 5375..6155 of consensus"
repeat_region     36330..36635
                   /note="AluSg repeat: matches 2..307 of consensus"
repeat_region     36833..36896
                   /note="MER72 repeat: matches 541..606 of consensus"
repeat_region     37018..37351
                   /note="AluYb repeat: matches 1..299 of consensus"
repeat_region     37352..38311
                   /note="L1M4 repeat: matches 3727..4692 of consensus"
repeat_region     38312..38456
                   /note="AluY repeat: matches 123..269 of consensus"
repeat_region     38457..38542

```

```

repeat_region     /note="43 copies 2 mer aa 62% conserved"
repeat_region     /note="L1MD1 repeat: matches 5354..6224 of consensus"
repeat_region     39479..39699
                   /note="AluY repeat: matches 84..303 of consensus"
repeat_region     39928..40268
                   /note="L1M1 repeat: matches 5609..5957 of consensus"
repeat_region     40269..40548
                   /note="AluX repeat: matches 1..277 of consensus"
repeat_region     40829..40963
                   /note="MIR repeat: matches 94..248 of consensus"

Query Match      2.1%; Score 40.6; DB 9; Length 168108;
Best Local Similarity 57.5%; Pred.No.5;
Matches 73; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 61 aaacattgaaatactgctgagagagtgacatgatgactatgagactatcctaac 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 91971 AAAGATATAGATCATTCGCCAGTACCAATAGATCCCTCTGCTCTCATACACAR 91912

QY 121 tctcttgacatttatgtgtagtctccattctctgctgtaattccctgagtc 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 91911 CCTCTTTTTCCTCTATCTGTATATCTTGTGATTTTCTTTCTTATCTGTGTG 91852

QY 181 tgtagt 187
    || |||
DB 91851 TGCATGT 91845

```

```

RESULT 12
AL513187
LOCUS             122032 bp    DNA    linear    PRI 11-AUG-2001
DEFINITION       Human DNA sequence from clone RP11-436h6 on chromosome 1, complete
sequence.
ACCESSION        AL513187
VERSION          AL513187.6    GI:15149851
KEYWORDS         HTG.
SOURCE           human.
ORGANISM         Homo sapiens
                  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE        1 (bases 1 to 122032)
AUTHORS          Wallis,J.
TITLE            Direct Submission
JOURNAL          Submitted (11-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
                  CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
                  requests: clonequest@sanger.ac.uk
                  On Aug 13, 2001 this sequence version replaced gi:14160999.
                  During sequence assembly data is compared from overlapping clones.
                  Where differences are found these are annotated as variations
                  together with a note of the overlapping clone name. Note that the
                  variation annotation may not be found in the sequence submission
                  corresponding to the overlapping clone, as we submit sequences with
                  only a small overlap as described above.
                  This sequence was finished as follows unless otherwise noted: all
                  regions were either double-stranded or sequenced with an alternate
                  chemistry or covered by high quality data (i.e., phred quality >=
                  30); an attempt was made to resolve all sequencing problems, such
                  as compressions and repeats; all regions were covered by at least
                  one plasmid subclone or more than one M13 subclone; and the
                  assembly was confirmed by restriction digest. The following
                  abbreviations are used to associate primary accession numbers given
                  in the feature table with their source databases: Em, EMBL; Sw,
                  SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
                  database can be found at
                  http://www.sanger.ac.uk/Projects/C_elegans/wormpep
                  This sequence
                  was generated from part of bacterial clone contigs of human
                  chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
                  Group. Further information can be found at
                  http://www.sanger.ac.uk/HGP/Chrl
                  RP11-436h6 is from the library RPc1-11.2 constructed by the group
                  of Pieter de Jong. For further details see
                  http://www.chori.org/bacpac/home.htm

```

```

COMMENT
repeat_region     /note="43 copies 2 mer aa 62% conserved"
repeat_region     /note="L1MD1 repeat: matches 5354..6224 of consensus"
repeat_region     39479..39699
                   /note="AluY repeat: matches 84..303 of consensus"
repeat_region     39928..40268
                   /note="L1M1 repeat: matches 5609..5957 of consensus"
repeat_region     40269..40548
                   /note="AluX repeat: matches 1..277 of consensus"
repeat_region     40829..40963
                   /note="MIR repeat: matches 94..248 of consensus"

Query Match      2.1%; Score 40.6; DB 9; Length 168108;
Best Local Similarity 57.5%; Pred.No.5;
Matches 73; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 61 aaacattgaaatactgctgagagagtgacatgatgactatgagactatcctaac 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 91971 AAAGATATAGATCATTCGCCAGTACCAATAGATCCCTCTGCTCTCATACACAR 91912

QY 121 tctcttgacatttatgtgtagtctccattctctgctgtaattccctgagtc 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 91911 CCTCTTTTTCCTCTATCTGTATATCTTGTGATTTTCTTTCTTATCTGTGTG 91852

QY 181 tgtagt 187
    || |||
DB 91851 TGCATGT 91845

```

VECTOR: PBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-436H6 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-436H6 is at 1 in this sequence. The true left end of clone RP11-270C12 is at 120033 in this sequence.

FEATURES
source

1. 122032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-436H6"
/clone_lib="RPC1-11.2"
BASE COUNT 37381 a 23078 c 22279 g 39294 t
ORIGIN

Query Match 2.1%; Score 40.4; DB 9; Length 122032;
Best Local Similarity 52.4%; Pred. No. 5.6;
Matches 89; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 18 aaggtcattgagatgtgttataatgtcaaggtccctatccaacattgaactgt 77
DB 3188 AATGACCATATTAAGAGCTCTATTAAGAGATGCTATTAAGTGAATATTT 3247
QY 78 ctagaaggagttgaacatgatcatatggacatcatcattccttggactttat 137
DB 3248 ACTGACGATTAAGAGTGTATCTTGAATGCTCACTTAAAGTATTTTCTTG 3307
QY 138 tgtgtagtctccattctctctgttgtaattccctcagagctgtgaagt 187
DB 3308 TGTATATTTCCAGATTTCTTACTTTCATGATTTTAAAGTATATGTGT 3357

RESULT 13
AL607133 161067 bp DNA linear HTG 15-NOV-2001
LOCUS Homo sapiens chromosome 1 clone RP11-157D21, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL607133.4 GI:16973213
VERSION HTG; HTGS_PHASE2; HTGS_CANCELLED.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Hall, R.
AUTHORS Submitted (14-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
TITLE Cambridge, CB10 1SA, UK. E-mail enquiries:
JOURNAL humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Nov 16, 2001 this sequence version replaced g1:16415681.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA102H4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 160914 bases at least Q40
Consensus quality: 160984 bases at least Q30
Consensus quality: 161024 bases at least Q20
Insert size: 161067; sum-of-contrigs
Insert size: 172236; 5.1% error; agarose-fp
Quality coverage: 10.07x in Q20 bases; sum-of-contrigs quality
coverage: 9.42x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 161067
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-157D21"
/clone_lib="RPC1-11.1"
misc_feature
/note="assembly-fragment:04132"

BASE COUNT 52153 a 29269 c 30412 g 49233 t
ORIGIN

Query Match 2.1%; Score 40.4; DB 2; Length 161067;
Best Local Similarity 52.4%; Pred. No. 5.7;
Matches 89; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 18 aaggtcattgagatgtgttataatgtcaaggtccctatccaacattgaactgt 77
DB 119992 AATGACCATATTAAGAGCTCTATTAAGAGATGCTATTAAGTGAATATTT 119933
QY 78 ctagaaggagttgaacatgatcatatggacatcatcattccttggactttat 137
DB 119932 ACTGACGATTAAGAGTGTATCTTGAATGCTCACTTAAAGTATTTTCTTG 119873
QY 138 tgtgtagtctccattctctctgttgtaattccctcagagctgtgaagt 187
DB 119872 TGTATATTTCCAGATTTCTTACTTTCATGATTTTAAAGTATATGTGT 119823

RESULT 14
AL590427 213692 bp DNA linear HTG 01-FEB-2002
LOCUS Homo sapiens chromosome 1 clone RP11-396N10, *** SEQUENCING IN
DEFINITION PROGRESS ***, 9 unordered pieces.
ACCESSION AL590427.12 GI:18491341
VERSION HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULFILLTOP.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Wallis, J.
AUTHORS Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
TITLE Cambridge, CB10 1SA, UK. E-mail enquiries:
JOURNAL humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Feb 4, 2002 this sequence version replaced g1:17902916.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA396N10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 0% of reads
Sequencing vector: plasmid; L08752; 99% of reads
Chemistry: Dye-terminator ET-amersham; 0% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Consensus quality: 210220 bases at least Q40
Consensus quality: 211827 bases at least Q30
Consensus quality: 212509 bases at least Q20
Insert size: 212892; sum-of-contrigs
Insert size: 188062; 4.0% error; agarose-fp
Quality coverage: 9.90x in Q20 bases; sum-of-contrigs quality
coverage: 13.68x in Q20 bases; agarose-fp

Tue Aug 20 08:07:59 2002

us-09-662-812-1_1.rge

Page 18

Oy 1696 ag 1697
Db 221 RS 220

Search completed: August 20, 2002, 01:14:17
Job time: 6607 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 23:24:10 ; Search time 2367.23 Seconds
(without alignments)
16858.060 Million cell updates/sec

Title: US-09-662-812-1

Perfect score: 1907

Sequence: 1 gtggcttgatttgaaaaag.....ctcatcaagagctgtgaaa 1907

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

- Listing first 45 summaries
- 1: gb_ba:*
 - 2: gb_htg:*
 - 3: gb_in:*
 - 4: gb_cm:*
 - 5: gb_ov:*
 - 6: gb_pat:*
 - 7: gb_ph:*
 - 8: gb_pl:*
 - 9: gb_pr:*
 - 10: gb_ro:*
 - 11: gb_sts:*
 - 12: gb_sy:*
 - 13: gb_un:*
 - 14: gb_vl:*
 - 15: em_ba:*
 - 16: em_fun:*
 - 17: em_hum:*
 - 18: em_in:*
 - 19: em_mu:*
 - 20: em_om:*
 - 21: em_or:*
 - 22: em_ov:*
 - 23: em_pat:*
 - 24: em_ph:*
 - 25: em_pl:*
 - 26: em_ro:*
 - 27: em_sts:*
 - 28: em_un:*
 - 29: em_vl:*
 - 30: em_htg_hum:*
 - 31: em_htg_inv:*
 - 32: em_htg_other:*
 - 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Query Score Match Length DB ID Description

1	1907	100.0	1907	6	AX100528	Sequence
2	1907	100.0	8268	1	AE002234	Chlamydia
3	1907	100.0	16448	1	AE001587	Chlamydia
4	1707	89.5	1707	6	AX349527	Sequence
5	561.4	29.4	11627	1	AE001308	Chlamydia
6	511.8	26.8	11765	1	AE002331	Chlamydia
7	84	4.4	2088	6	AX349569	Sequence
8	43	2.3	7892	6	AX348443	Sequence
9	41.6	2.2	1141	6	AP004050	Sequence
10	40.6	2.1	128866	9	AL445687	Human DNA
11	40.6	2.1	168108	9	AL513187	Human DNA
12	40.4	2.1	122032	9	AL513187	Human DNA
13	40.4	2.1	161067	2	AL607133	Homo sapi
14	40.4	2.1	213692	2	AL590427	Homo sapi
15	40	2.1	1141	6	AX083744	Sequence
16	39.6	2.1	22601	8	AB022222	Arabidops
17	39.6	2.1	69988	2	AC100581	Mus muscu
18	39.4	2.1	54585	2	AC100191	Mus muscu
19	39.4	2.1	59564	2	AC102477	Mus muscu
20	39.4	2.1	62396	2	AC100906	Mus muscu
21	39.4	2.1	165127	2	AC103172	Rattus no
22	39.2	2.1	1976	6	AX136235	Sequence
23	39.2	2.1	2311	9	AK027438	Homo sapi
24	39.2	2.1	2870	9	BC020959	Homo sapi
25	39.2	2.1	3389	9	AF160213	Homo sapi
26	39.2	2.1	3592	9	AF269150	Homo sapi
27	39.2	2.1	145593	9	AC069523	Homo sapi
28	39.2	2.1	157943	9	AC016938	Homo sapi
29	39.2	2.1	162820	9	AL138765	Human DNA
30	39	2.0	114314	9	AL359494	Human DNA
31	39	2.0	143840	2	AC091542	Felis cat
32	39	2.0	166602	2	AC022899	Homo sapi
33	39	2.0	169792	2	AC091508	Felis cat
34	39	2.0	180832	2	AC044828	Homo sapi
35	39	2.0	212310	2	AC008679	Homo sapi
36	38.8	2.0	24186	2	AC108968	Rattus no
37	38.6	2.0	92250	9	AC026693	Homo sapi
38	38.6	2.0	125020	9	AF429315	Homo sapi
39	38.6	2.0	156248	9	AC027322	Homo sapi
40	38.6	2.0	235564	2	AC022062	Mus muscu
41	38.4	2.0	11323	3	AF067951	Caenorhab
42	38.4	2.0	104972	9	AC025176	Homo sapi
43	38.4	2.0	152862	2	OSJN00168	Oryza sat
44	38.4	2.0	177838	2	AL606690	Oryza sat
45	38.4	2.0	178113	2	AC099516	Homo sapi

ALIGNMENTS

RESULT 1
AX100528 LOCUS AX100528 1907 bp DNA linear PAT 10-APR-2001
DEFINITION Sequence 1 from Patent WO0121804.
ACCESSION AX100528
VERSION AX100528.1 GI:13619532
KEYWORDS
SOURCE
ORGANISM Chlamydia pneumoniae.
REFERENCE Chlamydia pneumoniae
AUTHORS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
TITLE I chlamydia /i antigens and corresponding dna fragments and uses
JOURNAL Mardin, A.D., Oomen, R.P., Wang, J. and Dunn, P.
Patent: WO 0121804-A 1 29-MAR-2001;
FEATURES
source
1. 1907
location/Qualifiers
/organism="Chlamydia pneumoniae"
/db_xref="taxon:83558"
101. 1807
/note="unnamed protein product"
/codon_start=1

```
/transl_table=1
/protein_18="CAC36452.1"
/db_xref="GI:13619533"
/translation="MGLFHLTLFGLILCSLPISLVAKPESVGHKILYISTQOAL
ATYELADYADHDHDFVFLKIGEDYLKOSIHSSPPKRSITIGAGLAGSEALDVS
QAMEFADPDLQDLVLISAVSHGLKTSDDLKALISPPVRLAARLAKMKRYI
DHLSPFHKLPEEIOCLSAATFLRLEPESNAVIRDLIAKKSIRSTALQIEYOO
KRRLPTLRNLITSAPDOEALILVALGKLKDGOSYINIKOLOPDDVTLAAQALI
ALGKEEDALPVIRKQALEERPRALYLRHPLEIGIPALPFLKTNSEKLVNVAL
LLEIGCDTPKLEYITRERLYOPHYNETLALSFSGRTLOMKRVNIYVPODEREL
LSTTRGEEQILFRLPKFAYLPCYKLLASQKLTATTAISLSTSHOEDLL
FOAKLPGEPIIRAYADLAIVNLKDEPKRSLLDVAKPKLIOELFLVDPEMOPHS
MPYLRVYTPESRFRKLMIDILEFATSKSSDITLLIQLMBSGAKNFPVLAGLILKI
VE"
BASE COUNT      542 a      456 c      380 g      529 t
ORIGIN
Query Match      100.0%; Score 1907; DB 6; Length 1907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtggctgatttgaagaaagtcacatgagatggttataatgtaagtcctccatcc 60
Db 1 GTGGCTGATTTTGAAAAAGTCCATGATGTTTATATATGTTCAAGTCTCCCTATCC 60
QY 61 aaacatgaacttctgtagagagatggaacatgacatctatgagactatccatcaac 120
Db 61 AAACATGAAATCTTGTCTAGAGAGTGAACATGATCTATGGACATTTCCATCTAAC 120
QY 121 tctcttggaactttatgtagtctccatctctcttcttcttgaattccctgaatc 180
Db 121 TCTCTTGAGACTTTTATTTGTAGTCTTCCATTTCTCTTGTGTAATTCCTCGAGTTC 180
QY 181 ttagagtaaaagatcccttataaagtaacatctacacagagccttagacaata 240
Db 181 TGTAGGTATATAGATCTTTATATAGTACGCAATCTACACAGAGCGCTTAGCAACTA 240
QY 241 tctggaagctttagatcctcagtgatcatgactctctgctttaaagaaatcaga 300
Db 241 TCTGGAACCTTATAGTGGCTGAGTGCATGACTTCTTCTTTAGAAAAATCGGAGA 300
QY 301 agactatccaagcaaacatccactcctcagatccgcaactagaagaaacacatcat 360
Db 301 AGACTATCTCAAGCAAAACATCCACTCTCTCAGATCCGCAAACTGAAAAACACCATCAT 360
QY 361 tggagcagccttgccgagatcttcagaagccttggacgtgtctcccaagatggaac 420
Db 361 TGGAGCAGGCTTGGCGGAGCTTCAAGAACCTTGGAGCTGCTCCCAAGCTATGGAAC 420
QY 421 tgcagacccctcagcagcagctactggtttatcggcagctcgaagacatcttggaaac 480
Db 421 TGCAGACCCCTCGAGAGCTACTGGTTTTATCGCAGTCTCAGAGACATCTTGGGAAAC 480
QY 481 ttccgaagactactggtttaagcttttagacatccctatctctcactcgttagaagc 540
Db 481 TTCTGAGAGACTTACTGTTTAAAGCTTTTAGCATCTCCCTATCTCTCATCCGCTTAGAGC 540
QY 541 cgcctatagacttgcataattgaagaaacataaagtcattgatactatcttcat 600
Db 541 CGCCTATAGACTTCTATTTGAAGAACACTAAGTCAATGATCATCTACTCTTTTCAT 600
QY 601 tcaataagctccgagaagaatcaatgacctatctgaggcaatattcctgaogcttgaagc 660
Db 601 TCATAAAGCTTCCGAAAAAATCCAAATGCTATGCGGCAATATTTCTACGCTTGGAGAC 660
QY 661 tgaagaactcgaatgctataatctgggagatctctgagctgcaagaaagcgagatcgag 720
Db 661 TGAAGATCTGATGCTTATATTCGGGATCTCTTAGCTGCAAGAAAGCGGATTCGAG 720
QY 721 tgcacagacttgcagatcaggaataacaaacaaacgcttcttcgacacttagaa 780
Db 721 TGCACAGACTTGTGAGATCGGAGAAATACCAAAACAAACGCTTCTTCCGACACTTAGGAA 780
/transl_table=1
/protein_18="CAC36452.1"
/db_xref="GI:13619533"
/translation="MGLFHLTLFGLILCSLPISLVAKPESVGHKILYISTQOAL
ATYELADYADHDHDFVFLKIGEDYLKOSIHSSPPKRSITIGAGLAGSEALDVS
QAMEFADPDLQDLVLISAVSHGLKTSDDLKALISPPVRLAARLAKMKRYI
DHLSPFHKLPEEIOCLSAATFLRLEPESNAVIRDLIAKKSIRSTALQIEYOO
KRRLPTLRNLITSAPDOEALILVALGKLKDGOSYINIKOLOPDDVTLAAQALI
ALGKEEDALPVIRKQALEERPRALYLRHPLEIGIPALPFLKTNSEKLVNVAL
LLEIGCDTPKLEYITRERLYOPHYNETLALSFSGRTLOMKRVNIYVPODEREL
LSTTRGEEQILFRLPKFAYLPCYKLLASQKLTATTAISLSTSHOEDLL
FOAKLPGEPIIRAYADLAIVNLKDEPKRSLLDVAKPKLIOELFLVDPEMOPHS
MPYLRVYTPESRFRKLMIDILEFATSKSSDITLLIQLMBSGAKNFPVLAGLILKI
VE"
BASE COUNT      542 a      456 c      380 g      529 t
ORIGIN
Query Match      100.0%; Score 1907; DB 6; Length 1907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtggctgatttgaagaaagtcacatgagatggttataatgtaagtcctccatcc 60
Db 1 GTGGCTGATTTTGAAAAAGTCCATGATGTTTATATATGTTCAAGTCTCCCTATCC 60
QY 61 aaacatgaacttctgtagagagatggaacatgacatctatgagactatccatcaac 120
Db 61 AAACATGAAATCTTGTCTAGAGAGTGAACATGATCTATGGACATTTCCATCTAAC 120
QY 121 tctcttggaactttatgtagtctccatctctcttcttcttgaattccctgaatc 180
Db 121 TCTCTTGAGACTTTTATTTGTAGTCTTCCATTTCTCTTGTGTAATTCCTCGAGTTC 180
QY 181 ttagagtaaaagatcccttataaagtaacatctacacagagccttagacaata 240
Db 181 TGTAGGTATATAGATCTTTATATAGTACGCAATCTACACAGAGCGCTTAGCAACTA 240
QY 241 tctggaagctttagatcctcagtgatcatgactctctgctttaaagaaatcaga 300
Db 241 TCTGGAACCTTATAGTGGCTGAGTGCATGACTTCTTCTTTAGAAAAATCGGAGA 300
QY 301 agactatccaagcaaacatccactcctcagatccgcaactagaagaaacacatcat 360
Db 301 AGACTATCTCAAGCAAAACATCCACTCTCTCAGATCCGCAAACTGAAAAACACCATCAT 360
QY 361 tggagcagccttgccgagatcttcagaagccttggacgtgtctcccaagatggaac 420
Db 361 TGGAGCAGGCTTGGCGGAGCTTCAAGAACCTTGGAGCTGCTCCCAAGCTATGGAAC 420
QY 421 tgcagacccctcagcagcagctactggtttatcggcagctcgaagacatcttggaaac 480
Db 421 TGCAGACCCCTCGAGAGCTACTGGTTTTATCGCAGTCTCAGAGACATCTTGGGAAAC 480
QY 481 ttccgaagactactggtttaagcttttagacatccctatctctcactcgttagaagc 540
Db 481 TTCTGAGAGACTTACTGTTTAAAGCTTTTAGCATCTCCCTATCTCTCATCCGCTTAGAGC 540
QY 541 cgcctatagacttgcataattgaagaaacataaagtcattgatactatcttcat 600
Db 541 CGCCTATAGACTTCTATTTGAAGAACACTAAGTCAATGATCATCTACTCTTTTCAT 600
QY 601 tcaataagctccgagaagaatcaatgacctatctgaggcaatattcctgaogcttgaagc 660
Db 601 TCATAAAGCTTCCGAAAAAATCCAAATGCTATGCGGCAATATTTCTACGCTTGGAGAC 660
QY 661 tgaagaactcgaatgctataatctgggagatctctgagctgcaagaaagcgagatcgag 720
Db 661 TGAAGATCTGATGCTTATATTCGGGATCTCTTAGCTGCAAGAAAGCGGATTCGAG 720
QY 721 tgcacagacttgcagatcaggaataacaaacaaacgcttcttcgacacttagaa 780
Db 721 TGCACAGACTTGTGAGATCGGAGAAATACCAAAACAAACGCTTCTTCCGACACTTAGGAA 780
QY 781 ttctgtaacgagtgctctcctcaagatcaagaagctatctttagttagggagct 840
Db 781 TTTGTAACGAGTGGGTCTCTCAAGATCAAGACGATATCTTTATGCTTAGGGAAGCT 840
QY 841 taaggaatggtcagagctactacataataaagaacatgcaagcctgtagtgyatg 900
Db 841 TAAGATGGTCAAGCTCTACTCAATATATAAAGCAATTTGCGAAGCCGATGTGATGT 900
QY 901 caatttagcagagctcaagctttaaagcttcttgggaaagaaagggcgctcccgct 960
Db 901 CACTTtagcagagctcaagctttaaagcttcttgggaaagaaagggcgctcccgct 960
QY 961 gataaaaaagcaagcacttgaagagcgagcctcgaagccttgtagcttcaagcatalc 1020
Db 961 GATAAAAAAGCAAGCACTTGAAGAGCGGCGCTGAGCCCTGATGCTTAGCGCATCTACC 1020
QY 1021 ctctgagatagggatctccgatctgcccctgagatattcctaaactaagaacggaagc 1080
Db 1021 CTCTGAGATAGGATTCGATTCGATTCGCGCCGATATTCCTAAAGCAACGCGAAGC 1080
QY 1081 caagttgaatgagctttagagctctctcttagagtgtagtgtagcacccttaactatg 1140
Db 1081 CAAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 atacattacgaaagccttgctcccaacacatataatgagacttagccttgaagcttctc 1200
Db 1141 ATACATTCGCAAAAGGCTTGTCACACACATTAATGATGATGATGATGATGATGATGAT 1200
QY 1201 taaagggcgatcttatacaaaatctgaaagcggtgaaactatagctccctaaagatccca 1260
Db 1201 TAAAGGGCGATCTTATCAAAATGGAAGCGGCTGAACATCAATGATGATGATGATGATGAT 1260
QY 1261 ggaagggaaaggttgctctcccaacacgaggtctttaaagagcagatccttagcttct 1320
Db 1261 GGAGGGAAAGGTTGCTCTCCACACACAGGCTTAAAGAGCGATGATGATGATGATGATGAT 1320
QY 1321 ctccgcctacctaagaagcttaccctccctgtaattataagcttcttgaggagcaaga 1380
Db 1321 CTCCGCTTACCTTAAAGAGCTTACCTCCCTGATTAATGATGATGATGATGATGATGATGAT 1380
QY 1381 aactcagcttgcagactagctgattctctttaaagtaacacccctcaacacgaagcctt 1440
Db 1381 AACTCAGCTTGCACACTACTGCGATTTCTTTTAAAGTCAACCTCAATGCAAGAAAGCCTT 1440
QY 1441 agatcactcttccaagcttggaagcctcctgggaagccttcatcccgagcttagcaga 1500
Db 1441 AGATCTACTTTTCAAGCTGCGGAAGCTTCTGGAAGAACTTATCATCCGCTTATGCAAGA 1500
QY 1501 tcttgctattataatctcaccaagaatcctgaaaaaaaacgcttctctcatalgatalgc 1560
Db 1501 TCTTGCTATTATATATCTCACCAAGATCTGAAAAAAGCTTCTCTCATGATTAATGCT 1560
QY 1561 aaaaaagctaatcagaagaaacttgtaattgtgagcagggaaacaaacaaaccccatcc 1620
Db 1561 AAAAAAGCTAATTCAGGAACCTTGTATTTGTGTGACACGGAAAAACCAAAACCCCATCC 1620
QY 1621 cagcatgccctatctacgttatacagtgcaaccccaagaagcgttagaagccttagtga 1680
Db 1621 CAGCATGCCCTATCTAGCTTATCAAGTTCACCCCAAGAAAGCCGTAAGAGCTCATGTTGGA 1680
QY 1681 tatctagagaacttagccacactcgaagctctccgaagatataccggttatgtatacaact 1740
Db 1681 TATCTAGAGACACTAGCACACCTGGAAGTCTCGAAGATATCGTATGATGATGATGATGAT 1740
QY 1741 gatgacggaaggaatgcaaaaaaatctccagctccttgagagcttactcataaaaaatgtc 1800
Db 1741 GATGACGGAAGGAGATGCAAAAAATTTTCCAGTCTTGCAAGCTTACTCATTAATAAATTTGT 1800
QY 1801 ggaatlaaccccaactagcttctttagaagcttgcttctttagcttcttcttcttcttct 1860
Db 1801 GGAATTAACCCCACTTACGCTTATGAAAGCTTCTTATTTCTTACTGCTTCTTCTTCTTCT 1860
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 23:59:25 ; Search time 226.66 Seconds
(without alignments)
14445.230 Million cell updates/sec

Title: US-09-662-812-1
Perfect score: 1907
Sequence: 1 gtggctgatttgaaaag.....ctcatcaagagcgtgtgaa 1907

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1907	100.0	1907	22	AAD03024
2	39.2	2.1	1867	22	AAH46935
3	39.2	2.1	1976	22	AAK93822
4	39.2	2.1	2299	22	AAK94454
5	39.2	2.1	2311	22	AAH13994
6	39.2	2.1	2864	22	AAK94275
7	39.2	2.1	3076	22	AAH46951
8	39.2	2.1	3370	21	AAH16630
9	38.2	2.0	38771	22	AAK66361

C	10	38.2	2.0	38771	22	AAK68883	Human immune/haema
C	11	38.2	2.0	38771	22	AAK69604	Human immune/haema
C	12	38.2	2.0	38771	22	AAK71499	Human immune/haema
C	13	38.2	2.0	38771	22	AAK72925	Human immune/haema
C	14	38.2	2.0	38771	22	AAK75870	Human immune/haema
C	15	38.2	2.0	38771	22	AAK80325	Human immune/haema
C	16	38.2	2.0	38771	22	AAK83538	Human immune/haema
C	17	37.2	2.0	932	21	AAK16080	Human prostate can
C	18	36.8	1.9	4940	16	AAO91286	ORC1 gene encoding
C	19	36.8	1.9	4940	18	AAK73279	S. cerevisiae orig
C	20	36.4	1.9	633	7	AAH60340	Sequence encoding
C	21	36	1.9	549	22	AAK44753	Human contig polyn
C	22	36	1.9	1180	22	AAH70004	Tomato glutamate d
C	23	36	1.9	2536	22	AAK64216	Human secreted pro
C	24	35.8	1.9	467	22	AAK93524	Human polynucleoti
C	25	35.6	1.9	336	22	AAH12496	Human cDNA clone (
C	26	35.6	1.9	2459	22	AAH16543	Human cDNA sequenc
C	27	35.4	1.9	21748	23	ABL02520	Drosophila melanog
C	28	35.2	1.8	482	22	AAK94479	Human neuroblastom
C	29	35.2	1.8	2379	22	AAH11120	Human polynucleoti
C	30	35.2	1.8	2387	23	ABL23408	Drosophila melanog
C	31	35.2	1.8	2413	22	AAK59334	Human polynucleoti
C	32	35.2	1.8	2498	21	AAK23778	Human fetal brain
C	33	35.2	1.8	3466	22	AAH16111	Human cDNA sequenc
C	34	35.2	1.8	4222	23	ABL02016	Drosophila melanog
C	35	35.2	1.8	4755	23	ABL01896	Drosophila melanog
C	36	35.2	1.8	6021	23	ABL14540	Drosophila melanog
C	37	35.2	1.8	6692	23	ABL14942	Drosophila melanog
C	38	35.2	1.8	6692	23	ABL20236	Drosophila melanog
C	39	35	1.8	4640	23	ABL08094	Drosophila melanog
C	40	35	1.8	633	7	AAH60341	Sequence encoding
C	41	34.8	1.8	1201	16	AAO98905	Coho Salmon growth
C	42	34.8	1.8	33962	21	AAH81528	N. meningitidis pa
C	43	34.8	1.8	34980	21	AAK21609	Neisseria meningit
C	44	34.8	1.8	34980	21	AAK21610	Neisseria meningit
C	45	34.8	1.8	34980	21	AAK21610	Neisseria meningit

ALIGNMENTS

RESULT 1	
AAD03024	
ID	AAD03024 standard; DNA; 1907 BP.
AC	AAD03024;
DT	13-JUN-2001 (first entry)
DE	Chlamydia pneumoniae outer membrane protein (OMP) gene.
KM	Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;
KW	antibody; medicament; Chlamydia infection; ds.
OS	Chlamydia pneumoniae.
FT	Key
FT	Location/Qualifiers
FT	101..1807
FT	/tag= a
FT	/product= "C. pneumoniae outer membrane protein (OMP)"
PD	WO200121804-A1.
PD	29-MAR-2001.
PF	15-SEP-2000; 2000WO-CA01088.
PR	20-SEP-1999; 99US-0154652.
PA	(AVET) AVENTIS PASTEUR LTD.
PI	Murdin AD, Coenen RP, Wang J, Dunn P;

DR WP1: 2001-244939/25.
DR P-PSDB: AAY72972.
XX
PT Novel Chlamydia pneumoniae outer membrane protein and polynucleotides
PT encoding them, useful as components of vaccines for treating Chlamydia
PT infections, and for detecting Chlamydia infections in the body fluids
PT of mammals.
XX
PS Claim 2; Fig 1; 82pp; English.
XX
CC The present sequence is a DNA encoding Chlamydia pneumoniae OMP (outer
CC membrane protein). OMP is useful in the production of vaccines,
CC antibodies and pharmaceutical compositions which are useful for treating
CC or preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or
CC C. pecorum) infections. They are also useful as diagnostic reagents for
CC detecting Chlamydia infection which involves assaying a body fluid
CC of a mammal to be tested for the components. The OMP vaccine is
CC useful in the preparation of a medicament for preventing and/or
CC treating Chlamydia infection. The primers derived from OMP gene are
CC also useful for detecting and/or identifying Chlamydia in a biological
CC material. OMP antibodies are also useful as reagents for purifying OMP
CC from a biological sample which involves carrying out antibody-based
CC affinity chromatography with the biological sample. OMP gene is
CC also useful in gene therapy.

XX
SQ Sequence 1907 BP; 542 A; 456 C; 380 G; 529 T; 0 other;

Query Match 100.0%; Score 1907; DB 22; Length 1907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcttgatttgaaaggctcagatgatttataatgctcaagcttcctacc 60
DB 1 gtgcttgatttgaaaggctcagatgatttataatgctcaagcttcctacc 60
QY 61 aaacttgaatacttgtagagaggtgaacatcgatctatggacatctcaac 120
DB 61 aaacttgaatacttgtagagaggtgaacatcgatctatggacatctcaac 120
QY 121 tctcttgagatttatgtgtatcttccattctcttctgtgttaattccctgagtc 180
DB 121 tctcttgagatttatgtgtatcttccattctcttctgtgttaattccctgagtc 180
QY 181 tgtatgcataagaatcctttatataagtaacgaatctacagaagagccttaagaacata 240
DB 181 tgtatgcataagaatcctttatataagtaacgaatctacagaagagccttaagaacata 240
QY 241 tctggaagcttagatgctacaggttgatcgaactcttcgttttaagaanaacgcgaga 300
DB 241 tctggaagcttagatgctacaggttgatcgaactcttcgttttaagaanaacgcgaga 300
QY 301 agactatctcaagaagaagcatcactcctcagatcgcgaacatagaaagacacatcat 360
DB 301 agactatctcaagaagaagcatcactcctcagatcgcgaacatagaaagacacatcat 360
QY 361 tggagcagcctgagcagatcttcagaagccttgagtgctctcccaagcttagaagac 420
DB 361 tggagcagcctgagcagatcttcagaagccttgagtgctctcccaagcttagaagac 420
QY 421 tggagcagcctgagcagatcttcagaagccttgagtgctctcccaagcttagaagac 480
DB 421 tggagcagcctgagcagatcttcagaagccttgagtgctctcccaagcttagaagac 480
QY 481 tcttgacgactactgtttaaagcttagatctccctatcctgtgtcaatccgcttagagc 540
DB 481 tcttgacgactactgtttaaagcttagatctccctatcctgtgtcaatccgcttagagc 540
QY 541 cgccttagagcttgtaatttgaagaacataagtcattgatcatctacatcttctcat 600
DB 541 cgccttagagcttgtaatttgaagaacataagtcattgatcatctacatcttctcat 600
QY 601 tcataagctcccgagaagaatccaatgacctatctgagcaatatctcctagccttgagac 660

DB 601 tcataagctcccgagaagaatccaatgacctatctgagcaatatctcctagccttgagac 660
QY 661 tgaagatctgtatgttatatctcgagatctcttagctgcgaagaagaagcagctcgag 720
DB 661 tgaagatctgtatgttatatctcgagatctcttagctgcgaagaagaagcagctcgag 720
QY 721 tgcacagcttgcagatcgagaatccacaacaaagcttcttcgcacattagaga 780
DB 721 tgcacagcttgcagatcgagaatccacaacaaagcttcttcgcacattagaga 780
QY 781 ttgttaacgaatgtgtctccctcaagaatcaagaacttctttagctttggagaagt 840
DB 781 ttgttaacgaatgtgtctccctcaagaatcaagaacttctttagctttggagaagt 840
QY 841 taagatgtgcagagctactacataataaaagaatgtcagaagcctgagtgtagt 900
DB 841 taagatgtgcagagctactacataataaaagaatgtcagaagcctgagtgtagt 900
QY 901 cactttagcagagctcaagcttataatgtcttgggagaagaagagcgtctccagt 960
DB 901 cactttagcagagctcaagcttataatgtcttgggagaagaagagcgtctccagt 960
QY 961 gataaaagcagaagcacttgagagcgctcgagccctgtatgcttagcgcatctacc 1020
DB 961 gataaaagcagaagcacttgagagcgctcgagccctgtatgcttagcgcatctacc 1020
QY 1021 ctctgagatagagatctcgatctgcctgcgatatctcctaaanaataagaacagagagc 1080
DB 1021 ctctgagatagagatctcgatctgcctgcgatatctcctaaanaataagaacagagagc 1080
QY 1081 caagttgaatgtagcttagctctcttagaggttaggtgtgacacccctaactctgta 1140
DB 1081 caagttgaatgtagcttagctctcttagaggttaggtgtgacacccctaactctgta 1140
QY 1141 atacattccgaaagcctgtgccacacatataatgaagccttagagcttgatcttc 1200
DB 1141 atacattccgaaagcctgtgccacacatataatgaagccttagagcttgatcttc 1200
QY 1201 taagggcgtaacttatacaanaattggaagcgggtgaacatcaatgcccaagaatcccca 1260
DB 1201 taagggcgtaacttatacaanaattggaagcgggtgaacatcaatgcccaagaatcccca 1260
QY 1261 ggaagagggaaggtgtgtctctccacaacccgaggtcttgaagaagcagatccttact 1320
DB 1261 ggaagagggaaggtgtgtctctccacaacccgaggtcttgaagaagcagatccttact 1320
QY 1321 ctccgcctactcaagaaggttactccctcctgtatattatagcttttgggaagtcaaga 1380
DB 1321 ctccgcctactcaagaaggttactccctcctgtatattatagcttttgggaagtcaaga 1380
QY 1381 aactcagcttgcactactgagatcttctttaaagtcacacccacatcagagaagcctt 1440
DB 1381 aactcagcttgcactactgagatcttctttaaagtcacacccacatcagagaagcctt 1440
QY 1441 agatctacttctcaagctcgaaagcttctctggaacatcatccgcgcttatgaga 1500
DB 1441 agatctacttctcaagctcgaaagcttctctggaacatcatccgcgcttatgaga 1500
QY 1501 tcttgctatttataatctcccaagaagctcctgaaaaaagcttctcctcatgtattgc 1560
DB 1501 tcttgctatttataatctcccaagaagctcctgaaaaaagcttctcctcatgtattgc 1560
QY 1561 aaaaaagctaatcagaagaacctgttatctgtggaacaggaagaacacacccatcc 1620
DB 1561 aaaaaagctaatcagaagaacctgttatctgtggaacaggaagaacacacccatcc 1620
QY 1621 cagcatgcccattactacgttatcaggtcaccgccgaagaagcgttagagaagctatgttga 1680
DB 1621 cagcatgcccattactacgttatcaggtcaccgccgaagaagcgttagagaagctatgttga 1680
QY 1681 tattctagaagacactgacacccctcgaaagcttctcgagaagatatccgttatgtataaact 1740

Db 1681 tatcttagagacactagcacctcgaaagctctccgaagatataccggttatgtatgaacaact 1740
Qy 1741 gatagcggaaggaatgcataaaatttcccaagctcttgcaagcttactctcataaaatgct 1800
Db 1741 gatagcggaaggaatgcataaaatttcccaagctcttgcaagcttactctcataaaatgct 1800
Qy 1801 ggaatacccccaacctgcgtttatgaaagcttgccttcttcttactcttcccttggctc 1860
Db 1801 ggaatacccccaacctgcgtttatgaaagcttgccttcttcttactcttcccttggctc 1860
Qy 1861 ttatgggtcctcagctgatacttggactcatacaagagcgtgtgaa 1907
Db 1861 ttatgggtcctcagctgatacttggactcatacaagagcgtgtgaa 1907
RESULT 2
AAH46935/c
ID AAH46935 standard; cDNA; 1867 BP.
XX
AC AAH46935;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human secreted protein encoding cDNA (clone id HB0EG11).
XX
KW Secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; virucide; fungicide; human;
KW ophthalmological; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200155430-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01431.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 12-SEP-2000; 2000US-0231968.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
PI Ni J, Ruden SM, Barash SC;
XX
WP1: 2001-476220/51.
DR P-PSDB; AAB85525.
XX
PT 17 Isolated nucleic acid molecules encoding human secreted proteins,
PT used to preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 414-415; 482pp; English.
XX
CC The invention provides novel human secreted proteins and polynucleotides
CC encoding them. The secreted proteins can be expressed by standard
CC recombinant methodology. The secreted proteins and polynucleotides are
CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
CC also be used in diagnosing a pathological condition. The antibodies to
CC the proteins can also be used in alleviating symptoms associated with the
CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
CC linked immunosorbent assays (ELISA). Disorders which are diagnosed or
CC treated include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for

CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotaxis. The polypeptides can also be used as a food additive or
CC preservative to increase or decrease storage capabilities. The present
CC sequence represents a human secreted protein encoding cDNA.
XX
SQ Sequence 1867 BP; 542 A; 315 C; 347 G; 663 T; 0 other;
Query Match 2.1%; Score 39.2; DB 22; Length 1867;
Best Local Similarity 47.9%; Pred. No. 0.62;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
Qy 14 gaaaaggtccatgagatgtgttataatgctcaagagtcctccatccatacaaatgaata 73
Db 1420 GAAAAAAGATATATCTACTTTAGACATTCAGTTAAATGTAGTTATCTTAATCTCAAAATG 1361
Qy 74 ctgctagaagagatgtaacatcgatcctatggagctatccatcctccttggactt 133
Db 1360 TTTAATAAACAAGATATCTCTCCATTTAACAATTGCTTCTTAACGTACAGTAAAT 1301
Qy 134 ttatgtgtagctctccatctctcttctgtctaataatccctgagctgtgataag 193
Db 1300 GCATTGTAGAGATACACTTCTGTCAAACTGTATCTTTGGATGGAATTAAATG 1241
Qy 194 atcccttataagtagcaatctacacagcagagccttagaacaatctgaaagc 249
Db 1240 TAACGTATAGTTTAAATTAATTAATGGAAGTTGGTCCACTTAAGATGACAGC 1185
RESULT 3
AAF93822/c
ID AAF93822 standard; cDNA; 1976 BP.
XX
AC AAF93822;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human cDNA encoding a membrane or secretory protein clone PSEC0143.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; ss.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PI 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
WP1: 2001-093989/11.
DR P-PSDB; AAB88395.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 1; SEQ ID 157; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences

CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantify the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
CC (ELISA)). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
CC
XX Sequence 1976 BP; 595 A; 314 C; 371 G; 696 T; 0 other;
SQ

Query Match 2.1%; Score 39.2; DB 22; Length 1976;
Best Local Similarity 47.9%; Pred. No. 0.65;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 14 gaaaaaggtccatgagatgtttataatgltcaaggctcccatccaacattgaata 73
DB 1231 GAAAAAAGATATACCTTTAGACATTCAATAAATGTATTAATCAAAATG 1172
QY 74 ctctgtagagagttgaacatcgatcctatgagcattccatccatcctcttgactt 133
DB 1171 TTTAATTAATAACAGTATCTTCTCCATTAAACCTTCTTCTTAACGTGACAGTAAT 1112
QY 134 ttatgtgtagtctccatctctctgtgtctaattccctgagctgtagtcataag 193
DB 1111 GCATTGTGAGAGTACACTTCTGCTTCAAACTGTAATCTTTTGATGAGATTAAGT 1052
QY 194 atcctttataaagtagcgaatctacacagcagccttagaacaatctggaagc 249
DB 1051 TAACTGTATAGTTTAAAGATAAATAATGGGAAGTTGGTCCAACTAAGATGACAGC 996

RESULT 4
AAK94454/C
ID AAK94454 standard; CDNA: 2299 BP.
XX
AC AAK94454;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 3256.
XX
NM Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PE 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI: 2001-524255/58.
XX
DR P-PSDB: AAM93524.
DR

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation.
XX
PS Claim 8; SEQ ID NO 3256; 1380bp + sequence listing; English.
XX

CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' and 3' ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 2299 BP; 649 A; 380 C; 457 G; 813 T; 0 other;
SQ

Query Match 2.1%; Score 39.2; DB 22; Length 2299;
Best Local Similarity 47.9%; Pred. No. 0.7;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 14 gaaaaaggtccatgagatgtttataatgltcaaggctcccatccaacattgaata 73
DB 1867 GAAAAAAGATATACCTTTAGACATTCAATAAATGTATTAATCAAAATG 1808
QY 74 ctctgtagagagttgaacatcgatcctatgagcattccatccatcctcttgactt 133
DB 1807 TTTAATTAATAACAGTATCTTCTCCATTAAACCTTCTTCTTAACGTGACAGTAAT 1748
QY 134 ttatgtgtagtctccatctctctgtgtctaattccctgagctgtagtcataag 193
DB 1747 GCATTGTGAGAGTACACTTCTGCTTCAAACTGTAATCTTTTGATGAGATTAAGT 1688
QY 194 atcctttataaagtagcgaatctacacagcagccttagaacaatctggaagc 249
DB 1687 TAACTGTATAGTTTAAAGATAAATAATGGGAAGTTGGTCCAACTAAGATGACAGC 1632

RESULT 5
AAH13994/C
ID AAH13994 standard; CDNA: 2311 BP.
XX
AC AAH13994;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11070.
XX
NM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

XX DR WPI; 2001-318749/34.
XX PT primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX PS Claim 8; SEQ ID 11070; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2311 BP; 655 A; 382 C; 458 G; 816 T; 0 other;

Query Match 2.1%; Score 39.2; DB 22; Length 2311;
Best Local Similarity 47.9%; Pred. No. 0.7;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 14 gaaaaggtccatgagatggtttataatgltcaaggtccctcatccaacatgaaata 73
DB 1879 GAAATAAGATATCATCTTTAGACATTCAGTTAAATGATGTTATCTTAATCTCAAAATG 1820
QY 74 ctgtcagagaggttgacatcgatcctatgagactatccatcaactcttggact 133
DB 1819 TTTAATATAAACAAGTATCTTCCATTTAACACTTTGCTTTCTAACTGTACAGTAATTT 1760
QY 134 ttattgttagtcttcccatctctgtgtgctaataatccctgagctgtgagtcataag 193
DB 1759 GCATTGTAGAGAGTACACTCTGTCTTCAAACTGTATCTTTGGATGGATTAATGATG 1700
QY 194 atccctataataagtcacgaactacacagcagcgcttagcaacatactcggaaagc 249
DB 1699 TAACTGTATAGTTTAAAGATTAATTAATGGGAAGTTGGTCCAACTAAGATGACAGC 1644

RESULT 6
AAK94275/c
ID AAK94275 standard; cDNA; 2864 BP.
XX AC AAK94275;
XX
XX DT 06-NOV-2001 (first entry)
XX DE Human full-length cDNA, SEQ ID NO: 2908.
XX DE Human full-length cDNA; cDNA synthesis; oligo-capping; ss.
XX OS Homo sapiens.
XX PN EP1130094-A2.
XX

PD 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-0114089.
XX
XX 08-JUL-1999; 99JP-0194486.
XX 11-JAN-2000; 2000JP-0118774.
XX 02-MAY-2000; 2000JP-0183765.
XX
XX PA (HELI-) HELIX RES INST.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI: 2001-524255/58.
XX P-PSDB; AAM93355.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
XX
XX PS Claim 8; SEQ ID NO 2908; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 2864 BP; 834 A; 463 C; 569 G; 998 T; 0 other;

Query Match 2.1%; Score 39.2; DB 22; Length 2864;
Best Local Similarity 47.9%; Pred. No. 0.8;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 14 gaaaaggtccatgagatggtttataatgltcaaggtccctcatccaacatgaaata 73
DB 2119 GAAATAAGATATCATCTTTAGACATTCAGTTAAATGATGTTATCTTAATCTCAAAATG 2060
QY 74 ctgtcagagaggttgacatcgatcctatgagactatccatcaactcttggact 133
DB 2059 TTTAATATAAACAAGTATCTTCCATTTAACACTTTGCTTTCTAACTGTACAGTAATTT 2000
QY 134 ttattgttagtcttcccatctctgtgtgctaataatccctgagctgtgagtcataag 193
DB 1999 GCATTGTAGAGAGTACACTCTGTCTTCAAACTGTATCTTTGGATGGATTAATGATG 1940
QY 194 atccctataataagtcacgaactacacagcagcgcttagcaacatactcggaaagc 249
DB 1939 TAACTGTATAGTTTAAAGATTAATTAATGGGAAGTTGGTCCAACTAAGATGACAGC 1884

RESULT 7
AAH46951/c
ID AAH46951 standard; cDNA; 3076 BP.
XX AC AAH46951;
XX
XX DT 25-SEP-2001 (first entry)
XX DE Human secreted protein encoding cDNA (clone id HBOEG11).
XX DE Secreted protein; immunosuppressive; antiarthritic; antirheumatic;
XX antiproliferative; cycostatic; cardiant; vasotropic; cerebroprotective;
XX neurotropic; neuroprotective; antibacterial; virucide; fungicide; human;
XX opthalmological; gene therapy; ss.
XX

OS Homo sapiens.
 FN WO200155430-A1.
 XX 02-AUG-2001.
 XX 17-JAN-2001; 2001MO-US01431.
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 12-SEP-2000; 2000US-0231968.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
 PI Ni J, Ruben SM, Barash SC;
 XX WPI: 2001-476220/51.
 DR P-PSDB: AAB85541.
 DR 17 isolated nucleic acid molecules encoding human secreted proteins,
 PT used to preventing, treating or ameliorating a medical condition -
 PS Claim 1; Page 426-427; 482pp: English.
 XX The invention provides novel human secreted proteins and polynucleotides
 CC encoding them. The secreted proteins can be expressed by standard
 CC recombinant methodology. The secreted proteins and polynucleotides are
 CC used to prevent, treat or ameliorate a medical condition in e.g. humans,
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
 CC also be used in diagnosing a pathological condition. The antibodies to
 CC the proteins can also be used in alleviating symptoms associated with the
 CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
 CC linked immunosorbent assays (ELISA). Disorders which are diagnosed or
 CC treated include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. The polypeptides can also be
 CC used to aid wound healing and epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities. The present
 CC sequence represents a human secreted protein encoding cDNA.
 XX
 XX Sequence 3076 BP; 940 A; 487 C; 617 G; 1028 T; 4 other;
 Query Match 2.1%; Score 39.2; DB 22; Length 3076;
 Best Local Similarity 47.9%; Pred. No. 0.83;
 Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 14 gaaaagagtcagatggttataatgtccaagctccctccaacatgaata 73
 DB 2370 GAAATAAAGATATACCTTACACATGTTAAATGATGTTATCTAAATCTCAAAATG 2511
 QY ctctctagagagtggaacatcgatctatggaattccatccaactcttggactt 133
 DB 2510 TTTAATAAATAACAGATCTCTTCATTTAACAATTGCTTTGTAACGTGACATAAAT 2451
 QY 134 ttattgttagtctccatctctctgttgtaaatccctgagctgtgagtcataag 193
 DB 2450 GCATTGTAGAGAGTACACTTCTGCTTCAAACTGATCTCTTTGATGGATTAAGT 2391
 QY 194 atccttataatagtaagcaatctacacagcaggccttagcaacatctggaagc 249
 DB 2390 TAACTGTATAGTTTAAAGATAAATAATGGAAGCTGTGTCACACTAAGATGACAGC 2335

RESULT 8

AAA1630/c
 ID AAA1630 standard; cDNA: 3370 BP.
 XX
 AC AAA1630;
 XX 16-JUN-2000 (first entry)
 DT
 XX Human secreted protein clone pk366_7 nucleotide sequence SEQ ID NO:25.
 DE
 XX Human; secreted protein; immunostimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytoskeletal; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoicide;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host disease;
 KW autoimmune inflammatory eye disease; allergy; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200009552-A1.
 XX 24-FEB-2000.
 PD
 XX 13-AUG-1999; 99WO-US18298.
 XX 14-AUG-1998; 98US-0096622.
 PR 17-AUG-1998; 98US-0096815.
 PR 04-SEP-1998; 98US-0099229.
 PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.
 XX
 XX (GENM) GENETICS INST INC.
 PA Jacobs K, McCoy JM, LaValle ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fechtel K;
 XX WPI: 2000-205979/18.
 DR P-PSDB: AAY94910.
 XX
 PT New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity -
 PS Claim 34; Page 493; 641pp: English.
 XX
 CC AAA1630 to AAA1637 encode the human secreted proteins given in
 CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,

CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AA16698 to AA16774 represent
CC probes for the human secreted proteins from the present invention.
XX

SQ Sequence 3370 BP; 1031 A; 537 C; 676 G; 1126 T; 0 other;

Query Match 2.1%; Score 39.2; DB 21; Length 3370;
Best Local Similarity 47.9%; Pred. No. 0.87;
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 14 gaaaaggtccatgaggtgtgtttataatgttcaagtcctccatccaacattgaata 73

DB 2571 GAAAAAGATATACATCTTAGACATTCAGTTAAATGTAATTATCTAAATCCAAATG 2512

QY 74 ctgtctagagaggttgaacatctatctatggaacttccatctcaactcttggactt 133

DB 2511 TTTAATTAATAACAAGTATCTCTCCATTTTAACTTTGCTTCTAAGTACAGTAAT 2452

QY 134 ttatitgttagtcttccatcttcttctgttgcctcaaatccctgagctgtgagtcataag 193

DB 2451 GCATTTGATAGAGTACACTTCTCTGTAACGTATCTTGTGATGGAATTAAATG 2392

QY 194 atcccttataatgacagcaatctacacagcagcccttagcaacatctggaagc 249

DB 2391 TAAGTATATGTTTAAAGATTAATAATGGAAGTTGTTCAACTAAGATGACAGC 2336

RESULT 9

AAK66361/c
ID AAK66361 standard; DNA; 38771 BP.

XX AAK66361;

XX 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21173.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytosolic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001MO-US01354.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225272.
PR 14-AUG-2000; 2000US-0225273.
PR 14-AUG-2000; 2000US-0225274.
PR 14-AUG-2000; 2000US-0225275.
PR 14-AUG-2000; 2000US-0225276.
PR 14-AUG-2000; 2000US-0225277.
PR 14-AUG-2000; 2000US-0225278.
PR 14-AUG-2000; 2000US-0225279.
PR 14-AUG-2000; 2000US-0225280.
PR 14-AUG-2000; 2000US-0225281.
PR 14-AUG-2000; 2000US-0225282.
PR 14-AUG-2000; 2000US-0225283.
PR 14-AUG-2000; 2000US-0225284.
PR 14-AUG-2000; 2000US-0225285.
PR 14-AUG-2000; 2000US-0225286.
PR 14-AUG-2000; 2000US-0225287.
PR 14-AUG-2000; 2000US-0225288.
PR 14-AUG-2000; 2000US-0225289.
PR 14-AUG-2000; 2000US-0225290.
PR 14-AUG-2000; 2000US-0225291.
PR 14-AUG-2000; 2000US-0225292.
PR 14-AUG-2000; 2000US-0225293.
PR 14-AUG-2000; 2000US-0225294.
PR 14-AUG-2000; 2000US-0225295.
PR 14-AUG-2000; 2000US-0225296.
PR 14-AUG-2000; 2000US-0225297.
PR 14-AUG-2000; 2000US-0225298.
PR 14-AUG-2000; 2000US-0225299.
PR 14-AUG-2000; 2000US-0225300.
PR 14-AUG-2000; 2000US-0225301.
PR 14-AUG-2000; 2000US-0225302.
PR 14-AUG-2000; 2000US-0225303.
PR 14-AUG-2000; 2000US-0225304.
PR 14-AUG-2000; 2000US-0225305.
PR 14-AUG-2000; 2000US-0225306.
PR 14-AUG-2000; 2000US-0225307.
PR 14-AUG-2000; 2000US-0225308.
PR 14-AUG-2000; 2000US-0225309.
PR 14-AUG-2000; 2000US-0225310.
PR 14-AUG-2000; 2000US-0225311.
PR 14-AUG-2000; 2000US-0225312.
PR 14-AUG-2000; 2000US-0225313.
PR 14-AUG-2000; 2000US-0225314.
PR 14-AUG-2000; 2000US-0225315.
PR 14-AUG-2000; 2000US-0225316.
PR 14-AUG-2000; 2000US-0225317.
PR 14-AUG-2000; 2000US-0225318.
PR 14-AUG-2000; 2000US-0225319.
PR 14-AUG-2000; 2000US-0225320.
PR 14-AUG-2000; 2000US-0225321.
PR 14-AUG-2000; 2000US-0225322.
PR 14-AUG-2000; 2000US-0225323.
PR 14-AUG-2000; 2000US-0225324.
PR 14-AUG-2000; 2000US-0225325.
PR 14-AUG-2000; 2000US-0225326.
PR 14-AUG-2000; 2000US-0225327.
PR 14-AUG-2000; 2000US-0225328.
PR 14-AUG-2000; 2000US-0225329.
PR 14-AUG-2000; 2000US-0225330.
PR 14-AUG-2000; 2000US-0225331.
PR 14-AUG-2000; 2000US-0225332.
PR 14-AUG-2000; 2000US-0225333.
PR 14-AUG-2000; 2000US-0225334.
PR 14-AUG-2000; 2000US-0225335.
PR 14-AUG-2000; 2000US-0225336.
PR 14-AUG-2000; 2000US-0225337.
PR 14-AUG-2000; 2000US-0225338.
PR 14-AUG-2000; 2000US-0225339.
PR 14-AUG-2000; 2000US-0225340.
PR 14-AUG-2000; 2000US-0225341.
PR 14-AUG-2000; 2000US-0225342.
PR 14-AUG-2000; 2000US-0225343.
PR 14-AUG-2000; 2000US-0225344.
PR 14-AUG-2000; 2000US-0225345.
PR 14-AUG-2000; 2000US-0225346.
PR 14-AUG-2000; 2000US-0225347.
PR 14-AUG-2000; 2000US-0225348.
PR 14-AUG-2000; 2000US-0225349.
PR 14-AUG-2000; 2000US-0225350.
PR 14-AUG-2000; 2000US-0225351.
PR 14-AUG-2000; 2000US-0225352.
PR 14-AUG-2000; 2000US-0225353.
PR 14-AUG-2000; 2000US-0225354.
PR 14-AUG-2000; 2000US-0225355.
PR 14-AUG-2000; 2000US-0225356.
PR 14-AUG-2000; 2000US-0225357.
PR 14-AUG-2000; 2000US-0225358.
PR 14-AUG-2000; 2000US-0225359.
PR 14-AUG-2000; 2000US-0225360.
PR 14-AUG-2000; 2000US-0225361.
PR 14-AUG-2000; 2000US-0225362.
PR 14-AUG-2000; 2000US-0225363.
PR 14-AUG-2000; 2000US-0225364.
PR 14-AUG-2000; 2000US-0225365.
PR 14-AUG-2000; 2000US-0225366.
PR 14-AUG-2000; 2000US-0225367.
PR 14-AUG-2000; 2000US-0225368.
PR 14-AUG-2000; 2000US-0225369.
PR 14-AUG-2000; 2000US-0225370.
PR 14-AUG-2000; 2000US-0225371.
PR 14-AUG-2000; 2000US-0225372.
PR 14-AUG-2000; 2000US-0225373.
PR 14-AUG-2000; 2000US-0225374.
PR 14-AUG-2000; 2000US-0225375.
PR 14-AUG-2000; 2000US-0225376.
PR 14-AUG-2000; 2000US-0225377.
PR 14-AUG-2000; 2000US-0225378.
PR 14-AUG-2000; 2000US-0225379.
PR 14-AUG-2000; 2000US-0225380.
PR 14-AUG-2000; 2000US-0225381.
PR 14-AUG-2000; 2000US-0225382.
PR 14-AUG-2000; 2000US-0225383.
PR 14-AUG-2000; 2000US-0225384.
PR 14-AUG-2000; 2000US-0225385.
PR 14-AUG-2000; 2000US-0225386.
PR 14-AUG-2000; 2000US-0225387.
PR 14-AUG-2000; 2000US-0225388.
PR 14-AUG-2000; 2000US-0225389.
PR 14-AUG-2000; 2000US-0225390.
PR 14-AUG-2000; 2000US-0225391.
PR 14-AUG-2000; 2000US-0225392.
PR 14-AUG-2000; 2000US-0225393.
PR 14-AUG-2000; 2000US-0225394.
PR 14-AUG-2000; 2000US-0225395.
PR 14-AUG-2000; 2000US-0225396.
PR 14-AUG-2000; 2000US-0225397.
PR 14-AUG-2000; 2000US-0225398.
PR 14-AUG-2000; 2000US-0225399.
PR 14-AUG-2000; 2000US-0225400.
PR 14-AUG-2000; 2000US-0225401.
PR 14-AUG-2000; 2000US-0225402.
PR 14-AUG-2000; 2000US-0225403.
PR 14-AUG-2000; 2000US-0225404.
PR 14-AUG-2000; 2000US-0225405.
PR 14-AUG-2000; 2000US-0225406.
PR 14-AUG-2000; 2000US-0225407.
PR 14-AUG-2000; 2000US-0225408.
PR 14-AUG-2000; 2000US-0225409.
PR 14-AUG-2000; 2000US-0225410.
PR 14-AUG-2000; 2000US-0225411.
PR 14-AUG-2000; 2000US-0225412.
PR 14-AUG-2000; 2000US-0225413.
PR 14-AUG-2000; 2000US-0225414.
PR 14-AUG-2000; 2000US-0225415.
PR 14-AUG-2000; 2000US-0225416.
PR 14-AUG-2000; 2000US-0225417.
PR 14-AUG-2000; 2000US-0225418.
PR 14-AUG-2000; 2000US-0225419.
PR 14-AUG-2000; 2000US-0225420.
PR 14-AUG-2000; 2000US-0225421.
PR 14-AUG-2000; 2000US-0225422.
PR 14-AUG-2000; 2000US-0225423.
PR 14-AUG-2000; 2000US-0225424.
PR 14-AUG-2000; 2000US-0225425.
PR 14-AUG-2000; 2000US-0225426.
PR 14-AUG-2000; 2000US-0225427.
PR 14-AUG-2000; 2000US-0225428.
PR 14-AUG-2000; 2000US-0225429.
PR 14-AUG-2000; 2000US-0225430.
PR 14-AUG-2000; 2000US-0225431.
PR 14-AUG-2000; 2000US-0225432.
PR 14-AUG-2000; 2000US-0225433.
PR 14-AUG-2000; 2000US-0225434.
PR 14-AUG-2000; 2000US-0225435.
PR 14-AUG-2000; 2000US-0225436.
PR 14-AUG-2000; 2000US-0225437.
PR 14-AUG-2000; 2000US-0225438.
PR 14-AUG-2000; 2000US-0225439.
PR 14-AUG-2000; 2000US-0225440.
PR 14-AUG-2000; 2000US-0225441.
PR 14-AUG-2000; 2000US-0225442.
PR 14-AUG-2000; 2000US-0225443.
PR 14-AUG-2000; 2000US-0225444.
PR 14-AUG-2000; 2000US-0225445.
PR 14-AUG-2000; 2000US-0225446.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225448.
PR 14-AUG-2000; 2000US-0225449.
PR 14-AUG-2000; 2000US-0225450.
PR 14-AUG-2000; 2000US-0225451.
PR 14-AUG-2000; 2000US-0225452.
PR 14-AUG-2000; 2000US-0225453.
PR 14-AUG-2000; 2000US-0225454.
PR 14-AUG-2000; 2000US-0225455.
PR 14-AUG-2000; 2000US-0225456.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225458.
PR 14-AUG-2000; 2000US-0225459.
PR 14-AUG-2000; 2000US-0225460.
PR 14-AUG-2000; 2000US-0225461.
PR 14-AUG-2000; 2000US-0225462.
PR 14-AUG-2000; 2000US-0225463.
PR 14-AUG-2000; 2000US-0225464.
PR 14-AUG-2000; 2000US-0225465.
PR 14-AUG-2000; 2000US-0225466.
PR 14-AUG-2000; 2000US-0225467.
PR 14-AUG-2000; 2000US-0225468.
PR 14-AUG-2000; 2000US-0225469.
PR 14-AUG-2000; 2000US-0225470.
PR 14-AUG-2000; 2000US-0225471.
PR 14-AUG-2000; 2000US-0225472.
PR 14-AUG-2000; 2000US-0225473.
PR 14-AUG-2000; 2000US-0225474.
PR 14-AUG-2000; 2000US-0225475.
PR 14-AUG-2000; 2000US-0225476.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225478.
PR 14-AUG-2000; 2000US-0225479.
PR 14-AUG-2000; 2000US-0225480.
PR 14-AUG-2000; 2000US-0225481.
PR 14-AUG-2000; 2000US-0225482.
PR 14-AUG-2000; 2000US-0225483.
PR 14-AUG-2000; 2000US-0225484.
PR 14-AUG-2000; 2000US-0225485.
PR 14-AUG-2000; 2000US-0225486.
PR 14-AUG-2000; 2000US-0225487.
PR 14-AUG-2000; 2000US-0225488.
PR 14-AUG-2000; 2000US-0225489.
PR 14-AUG-2000; 2000US-0225490.
PR 14-AUG-2000; 2000US-0225491.
PR 14-AUG-2000; 2000US-0225492.
PR 14-AUG-2000; 2000US-0225493.
PR 14-AUG-2000; 2000US-0225494.
PR 14-AUG-2000; 2000US-0225495.
PR 14-AUG-2000; 2000US-0225496.
PR 14-AUG-2000; 2000US-0225497.
PR 14-AUG-2000; 2000US-0225498.
PR 14-AUG-2000; 2000US-0225499.
PR 14-AUG-2000; 2000US-0225500.

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -

Disclosure; SEQ ID NO 21173; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patients own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting
the nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/haematopoietic-related diseases, especially
cancers and cancer metastases of haematopoietic-derived cells. AAK64703
to AAK87694 represent human immune/haematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAM82169
represent sequences used in the exemplification of the present invention.

Query Match 2.0%; Score 38.2; DB 22; Length 38771;
Best Local Similarity 50.8%; Pred. No. 7;
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 16 aaaagtcacatgcatgtgttataatgttcaaggtctccctacccaacatggaatact 75
DB 2840 AAATGATATATCATCTGTGGTTTGAATTCGATTCCTAATGACTAAAGATATTAACAG 2781
QY 76 tgctagaggaattgacacatcgatcgaacttgcacttccatcactccttggacttt 135
DB 2780 CTTTCACGAGCTTATTTGGCCATGATATGTCCTTTAGAGAGCTGTATTCATTC 2721
QY 136 atgtgtatgtctcccatctctctgtgtcctaattccctgagtcgtgtagtcataaga 194
DB 2720 ATTTGTAACTGGGTTATCTTTTATTTGTGTAATTCGAATTCAGAGCTTATGACACTGA 2662

RESULT 10
AAK68883/c
ID AAK68883 standard; DNA: 38771 BP.
XX
AC AAK68883;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23695.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX
KW Cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
PD 09-AUG-2001.
XX
FE 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249269.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251990.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruden SM;
XX WPI; 2001-483426/52.
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Disclosure; SEQ ID NO 23695; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 38771 BP; 10846 A; 8364 C; 8245 G; 9791 T; 1525 other;

Query Match 2.0%; Score 38.2; DB 22; Length 38771;
Best Local Similarity 50.8%; Pred. No. 7;
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 16 aaagtcacatgacgttcttaataatgttcaagctcccatcaacaacatgaact 75
DB 2840 AATATGATATATCATTTGCTGTTTGAATTTGCCATTTTCCCTAAAGCATTAAGATATTAAACAG 2781
QY 76 tgctagagaggttgacacatgatcatatgagactatccatcaactctcttgacttt 135

PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 24416; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SO Sequence 38771 BP; 9791 A; 8245 C; 8364 G; 10846 T; 1525 other;

Query Match 2.0%; Score 38.2; DB 22; Length 38771;
Best Local Similarity 50.8%; Pred. No. 7;
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

OY 16 aaagagtcgatgagtggttataatgtaagtcctccatccatacgaactgaataact 75
DB 35932 aaagaatatacatgttggttgcatttccttaagtaagtaagaatataacag 35991

OY 76 tgcataagagagtggaacatgatcatgagcatatccatcattcttggacttt 135
DB 35992 ctttcacgagctatgtgcacatgatagtctctctttagaagaactgtcatcatatcc 36051

OY 136 atgtgtgagtcctccatctctgtgtgtaaatccctgagctgtgagtcataaga 194
DB 36052 attttgaactgggtatattcttattgttgtaattctaagagttccttagacatcga 36110

RESULT 12
AAK71499/C
ID AAK71499 standard; DNA: 38771 BP.

XX
AC AAK71499;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26311.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 05-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.

PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPL; 2001-483426/52.	
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating cancers and	
PT	metastasis -	
XX		
PS	Disclosure; SEQ ID NO 26311; 3071pp + Sequence Listing; English.	
XX		
CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
CC	amino acid sequences given in AAK82170 to AAM91921. (I) have cytosolic	
CC	activity, and can be used in gene therapy and vaccine production. (II)	
CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (II) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (II), by inserting	
CC	the nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/haematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169	
CC	represent sequences used in the exemplification of the present invention.	
XX		
XX		
SQ	Sequence 38771 BP; 10846 A; 8364 C; 8245 G; 9791 T; 1525 other:	
Query Match 2.0%; Score 38.2; DB 22; Length 38771;		
Best Local Similarity 50.8%; Pred. No. 7;		
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;		
OY	16 aaaagtcacatgatgctgttataatgtccaaggctccctcatcacaacttgaaact 75	
DB	2840 AAATGATATATCATTTGTGGTATTTCATTTGCATTTCCCTAATGACATAAGATTAACAG 2781	
OY	76 tgcctagagaattgaacatcgatcctaaggcaattccatcataactctctttggacttt 135	
DB	2780 GTTTTCACAGACTTATTTGGCCATGATAATATGCTCTTTTAGAAGACTCTCATTCATATGCC 2721	
OY	136 attgtgttagcttcccattctctgtgtgtcctaaattccccctgagctgttagtcataga 194	
DB	2720 ATTGTGTAACCTGGGTATCTTTTATTGTGTAATTCATAAGAGTTCCTTTAGACATCTGA 2662	
RESULT 13		
ID	AAK72925	
ID	AAK72925 standard; DNA; 38771 BP.	
XX		
AC	AAK72925;	
XX		
DT	06-NOV-2001 (first entry)	
XX		
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27737.	
XX		
KW	Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;	
XX	cytostatic; gene therapy; vaccine; metastasis; ds.	
XX		
OS	Homo sapiens	

XX WO200157182-A2.
XX
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-483426/52.
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Disclosure: SEQ ID NO 27737; 3071bp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 38771 BP; 9791 A; 8245 C; 8364 G; 10846 T; 1525 other;

Query Match
Best Local Similarity 2.0%; Score 38.2; DB 22; Length 38771;
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 16 aaaaagtcagtgtgtgttataatgtccaaggtctcccatccaacatgaataact 75
Db 35932 aatgatatacatcattgtgttattgttcgtccctcccaagactaaagataataaag 35991

QY 76 tgcacagagagattgaacatgatataaggactatccatcaactctcttgactttt 135
Db 35992 ctttcacagacttatgtgcattatgtctcttttagaagactgtcattatcc 36051

QY 136 attgttagtctccatcttctgtgtcgaattccctcagctcgtgagtcataaga 194
Db 36052 atttgcataactggtgtatcttcttattgttgaattcgaagatccttagacatcga 36110

RESULT 14
AAK75870/C
ID AAK75870 standard; DNA: 38771 BP.
XX
XX AAK75870;
XX
DT 07-NOV-2001 (first entry)
XX
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:30682.
DE
XX
XX Human; immune; haematopoietic; immune/hematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX
XX W0200157182-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225216.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227093.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.

CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic
CC	activity, and can be used in gene therapy and vaccine production. (I)
CC	proteins and polynucleotides may be used in the prevention, diagnosis and
CC	treatment of diseases associated with inappropriate (I) expression. For
CC	example, they may be used to treat disorders associated with decreased
CC	expression by rectifying mutations or deletions in a patient's genome
CC	that affect the activity of (I) by expressing inactive proteins or to
CC	supplement the patients own production of (I). Additionally, (I)
CC	polynucleotides may be used to produce the secreted (I), by inserting
CC	the nucleic acids into a host cell and culturing the cell to express the
CC	protein. (I) proteins and polynucleotides may be used to prevent,
CC	diagnose and treat immune/haematopoietic-related diseases, especially
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC	to AAK87654 represent human immune/haematopoietic antigen genomic
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC	represent sequences used in the exemplification of the present invention.
CC	xx
SO	Sequence 38771 BP; 10846 A; 8364 C; 8245 G; 9791 T; 1525 other;
Query Match	2.0%; Score 38.2; DB 22; Length 38771;
Best Local Similarity	50.8%; Pred. No. 7;
Matches	91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY	16 aaaagtcacgagtgacatgcttataatgltcaaggtctccctacccaacattgaataact 75
Db	2840 AAATGATATATCATGTCGTGGTTTGAATTTGCATTTCCCTAATGACCTAATTAACAG 2781
QY	76 tgctagaagagtgacatgcacatgcatacgaagactatccatcctaactctcttgagcttt 135
Db	2780 CTTTTCACGACCTAATGTCATATGTCCTTTTAGAGACACGTCCTATTCATATCC 2721
QY	136 attgtgagtccttcccatctctctgtgtgcataatccctcgagctgtgagtcacaaga 194
Db	2720 ATTGTGTAACCTGGCTATCTTTTATGTGTGAATTCRAAGAGTCTTTAGACATCTGA 2662
RESULT 15	
AAK80325	
ID	AAK80325 standard; DNA; 38771 BP.
XX	
AC	AAK80325;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35137.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225256.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231245.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-483426/52.
 Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and
 metastasis -
 Disclosure; SEQ ID NO 35137; 3071bp + Sequence Listing; English.
 AA54951 to AA64702 encode the human immune/hematopoietic antigen (I)
 amino acid sequences given in AA62170 to AA61921. (I) have cytostatic
 activity, and can be used in gene therapy and vaccine production. (I)
 proteins and polynucleotides may be used in the prevention, diagnosis and
 treatment of diseases associated with inappropriate (I) expression. For
 example, they may be used to treat disorders associated with decreased
 expression by rectifying mutations or deletions in a patient's genome
 that affect the activity of (I) by expressing inactive proteins or to
 supplement the patients own production of (I). Additionally, (I)
 polynucleotides may be used to produce the secreted (I), by inserting
 the nucleic acids into a host cell and culturing the cell to express the
 protein. (I) proteins and polynucleotides may be used to prevent,

CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54542 to AAK54590 and
CC represent sequences used in the exemplification of the present invention.
XQ
XQ Sequence 38771 BP; 9791 A; 8245 C; 8364 G; 10846 T; 1525 other;
XQ

Query Match	2.0%;	Score 38.2;	DB 22;	Length 38771;
Best Local Similarity	50.8%;	Pred. No. 7;		
Matches 91; Conservative	0;	Mismatches 88;	Indels 0;	Gaps 0;

QY 16 aaaggtccagagatggtttataatgttcaaggtccoccatccaacaatggaatact 75
 Db 35932 aaagatatacatgttgggtttgtttgatttgcattccctaagtaagaagatattaaacag 35991
 QY 76 tgcataagagattgacatcgatcatatggagatcttcacatcaactcccttggacctt 135
 Db 35992 ctttcaacagacttatggccaatgtatatgtctcctttagaagaactgtctatcatcc 36051
 QY 136 atgtgtgattcttcccaattctctgtttgtgcaaaatcttcctgtgtaagtcataaga 194
 Db 36052 attttgtaactggttatccttttatgttgaattcaagaagttcctttaagacactga 36110

Search completed: August 20, 2002, 01:14:04
Job time: 4479 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 23:55:25 ; Search time 55.3 Seconds
(without alignments)
8470.580 Million cell updates/sec

Title: US-09-662-812-1
Perfect score: 1907
Sequence: 1 gtgctgattgattgaaaag.....ctcatcaagagcgtgtgaa 1907

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Dackfilest1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36.8	1.9	4940	US-08-484-105-1	Sequence 1, Appli
2	36.8	1.9	4940	US-08-484-106-1	Sequence 1, Appli
3	36	1.9	7218	US-08-232-463-14	Sequence 14, Appli
4	34.8	1.8	1201	US-08-286-872-1	Sequence 1, Appli
5	34.4	1.8	4301	US-08-121-446-3	Sequence 3, Appli
6	33.4	1.8	4706	US-08-331-081B-5	Sequence 5, Appli
7	33.4	1.8	4744	US-08-331-081B-3	Sequence 3, Appli
8	33.2	1.7	14602	US-08-597-236-1	Sequence 1, Appli
9	33.2	1.7	14602	US-08-746-682A-1	Sequence 1, Appli
10	33	1.7	767	US-08-998-416-472	Sequence 472, App
11	32.6	1.7	3621	US-09-019-201A-1	Sequence 1, Appli
12	32.6	1.7	31571	US-08-323-443B-1	Sequence 1, Appli
13	32.6	1.7	53526	US-08-658-136-2	Sequence 2, Appli
14	32.6	1.7	53577	US-08-658-136-1	Sequence 1, Appli
15	32.2	1.7	9636	US-08-323-170B-1	Sequence 1, Appli
16	32.2	1.7	9636	US-08-954-441-1	Sequence 1, Appli
17	32	1.7	4376	US-08-119-125A-1	Sequence 1, Appli
18	32	1.7	4380	US-07-582-945-1	Sequence 1, Appli
19	32	1.7	4380	US-08-453-141-1	Sequence 1, Appli
20	32	1.7	4380	US-08-293-314-1	Sequence 1, Appli
21	32	1.7	6744	US-08-119-125A-2	Sequence 2, Appli
22	31.8	1.7	6136	US-08-137-252-1	Sequence 1, Appli
23	31.8	1.7	30001	US-08-125-468-1	Sequence 1, Appli
24	31.8	1.7	30001	US-08-474-933-1	Sequence 1, Appli
25	31.2	1.6	1519	US-08-850-910A-44	Sequence 44, Appli
26	30.6	1.6	1275	US-08-793-666-9	Sequence 9, Appli
27	30.6	1.6	7218	US-08-232-463-14	Sequence 14, Appli

C 28	30.6	1.6	7505	4	US-09-078-294-13	Sequence 13, Appli
C 29	30.6	1.6	8600	4	US-09-457-708-1	Sequence 1, Appli
C 30	30.4	1.6	762	4	US-08-998-416-462	Sequence 462, App
C 31	30.2	1.6	1209	4	US-08-822-774-18	Sequence 18, Appl
C 32	30.2	1.6	1209	4	US-09-632-711-18	Sequence 18, Appl
C 33	30.2	1.6	2738	2	US-08-795-868-17	Sequence 17, Appl
C 34	30.2	1.6	2738	4	US-09-303-069-17	Sequence 17, Appl
C 35	30.2	1.6	10395	1	US-08-245-809-3	Sequence 3, Appli
C 36	30.2	1.6	10798	1	US-08-107-748-2	Sequence 2, Appli
C 37	30.2	1.6	10798	5	PCT-US82-01385-2	Sequence 2, Appli
C 38	30	1.6	1995	5	PCT-US96-05320A-1025	Sequence 1025, Ap
C 39	30	1.6	3055	1	US-08-236-754-1	Sequence 1, Appli
C 40	30	1.6	8655	3	US-09-075-272-1	Sequence 1, Appli
C 41	29.8	1.6	389	4	US-09-227-357-77	Sequence 77, Appl
C 42	29.8	1.6	2088	4	US-09-351-414-3	Sequence 3, Appli
C 43	29.6	1.6	960	2	US-08-245-511-3	Sequence 3, Appli
C 44	29.6	1.6	960	2	US-08-600-993A-3	Sequence 3, Appli
C 45	29.6	1.6	1999	3	US-08-961-083-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-484-105-1
; Sequence 1, Application US/08484105
; Patent No. 5589341
; GENERAL INFORMATION:
; APPLICANT: STILLMAN, Bruce
; APPLICANT: BELL, Stephen P
; APPLICANT: KOBAYASHI, Ryuji
; APPLICANT: RINE, Jasper
; APPLICANT: FOSS, Margit
; APPLICANT: MCNALLY, Francis J
; APPLICANT: LAURENSEN, Patricia
; APPLICANT: HERSKOWITZ, Ira
; APPLICANT: LI, Joachim J
; APPLICANT: GAVIN, Kimberly
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,105
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard Aron
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4940 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-484-105-1

Db 1401 yyyyyyyyyyyyyyyyyyyyyyygTACCAAATCTCTATCTCTTAAC 1460
QY 1548 tccatgatatgcaaaaagcctaatccagaaacctgttatcttgagacaggaacc 1607
Db 1461 tACTTGCAATAGATAGTAATTCACAGTGATGCCCTACATGCCGTTTTTGAACGTAAGA 1520
QY 1608 aaagaccacccacagcatgccctatctac 1637
Db 1521 TGGGTCTAGAGGATGCTACGCTACTACTAC 1550

RESULT 4

US-08-286-872-1/C
; Sequence 1, Application US/08286872
; Patent No. 5675061
; GENERAL INFORMATION:
; APPLICANT: Powers, Dennis A.
; TITLE OF INVENTION: Isolation and Characterization
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,872
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/192,272
; FILING DATE: 04-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0144.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Coho salmon Growth Hormone Gene
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 65..697
; OTHER INFORMATION: /note= "Growth Hormone Precursor"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 65..130
; OTHER INFORMATION: /note= "growth hormone signal"
; OTHER INFORMATION: peptide"
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 131..697
; OTHER INFORMATION: /note= "Growth Hormone"
US-08-286-872-1

Query Match 1.8%; Score 34.8; DB 1; Length 1201;

Best Local Similarity 57.3%; Pred. No. 0.67;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 1232 gtyaacatcatagtcctcaagatcccagagaggaaggtgtctctcacaccga 1291
Db 217 GTCAAAGTCATGTGACATTTCTGAGCCAAATAGTGATGTTGTCACCCGACTGACCGC 158
QY 1292 gftctgaagagagatcccttaagcttctcttcgcctactaagaagc 1341
Db 157 GATGTTGAAGAGCCGTTGTTTCTATGCTGCCCTTGACTGACGGAAC 108

RESULT 5

US-08-121-446-3
; Sequence 3, Application US/08121446
; Patent No. 6313276
; GENERAL INFORMATION:
; APPLICANT: IMURA, HIROO
; APPLICANT: NAKAO, KAZUMA
; APPLICANT: NAKANISHI, SHIGETADA
; TITLE OF INVENTION: A HUMAN ENDOTHELIN RECEPTOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,446
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/911,684
; FILING DATE: 10-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CIOTTI, THOMAS E.
; REGISTRATION NUMBER: 21,013
; REFERENCE/DOCKET NUMBER: 29900-20324.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 238..1566
US-08-121-446-3

Query Match 1.8%; Score 34.4; DB 4; Length 4301;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 86; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 797 tctccctcaagatcaagaagctatctttagctttaggaagctlaagatggtcagagc 856
Db 2431 TTTTTCAGGATATATTAATCTCTTTTTCATCTATCGTAGCTTAACCTGTTGGTT 2490
QY 857 tactacaataaaaaagcaatgcagagaccctgagtgtgagtcaacttaagcagagc 916
Db 2491 TTGTCACTGTGTAATACTTACCTACATACACATGATGATGATTAATAAGGAGGAGGC 2550


```
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/597,236
FILING DATE:
CLASSIFICATION: 426
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95201669.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci A., Allan
REGISTRATION NUMBER: 30256
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 14602 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 352..1803
OTHER INFORMATION: /product= "epsa"
FEATURE:
NAME/KEY: CDS
LOCATION: 1807..2535
OTHER INFORMATION: /product= "epsb"
FEATURE:
NAME/KEY: CDS
LOCATION: 2547..3239
OTHER INFORMATION: /product= "epsc"
FEATURE:
NAME/KEY: CDS
LOCATION: 3249..3995
OTHER INFORMATION: /product= "epsd"
FEATURE:
NAME/KEY: CDS
LOCATION: 4051..4731
OTHER INFORMATION: /product= "epse"
FEATURE:
NAME/KEY: CDS
LOCATION: 4898..5854
OTHER INFORMATION: /product= "epsf"
FEATURE:
NAME/KEY: CDS
LOCATION: 6425..7540
OTHER INFORMATION: /product= "epsg"
FEATURE:
NAME/KEY: CDS
LOCATION: 7736..8212
OTHER INFORMATION: /product= "epsh"
FEATURE:
NAME/KEY: CDS
LOCATION: 8221..9192
OTHER INFORMATION: /product= "epsi"
FEATURE:
NAME/KEY: CDS
LOCATION: 9285..10364
OTHER INFORMATION: /product= "epsj"
FEATURE:
NAME/KEY: CDS
LOCATION: 10392..11339
OTHER INFORMATION: /product= "epsk"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11302..12222
OTHER INFORMATION: /product= "CDS (eps L) covering CDS"
FEATURE:
NAME/KEY: CDS
LOCATION: 12233..13651
OTHER INFORMATION: /product= "epsm"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 352..1803
```

```
LOCATION: 13732..14305
OTHER INFORMATION: /function= "CDS on the
OTHER INFORMATION: complementary strand"
OTHER INFORMATION: /product= "orf2"
FEATURE:
NAME/KEY: terminator
LOCATION: 230..252
FEATURE:
NAME/KEY: promoter
LOCATION: 274..302
FEATURE:
NAME/KEY: RBS
LOCATION: 340..345
US-08-597-236-1

Query Match 1.78; Score 33.2; DB 1; Length 14602;
Best Local Similarity 57.8%; Pred. No. 9.2;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 1646 gtcccccagaagccgtacgaagctcatgttgatattctagagacactagccactcg 1705
Db 6131 GACCACTGCAAGCAGCAGCTCAATTACCTGGTGCATATTCGTAGTAACGTAAGTACCTAG 6072
Qy 1706 aagcttcggaagatccgtttatgtatatacaactgtagcg 1747
Db 6071 ACGTCCACCGAATATGACATTCTCTGTGTCTGCGCAAGACG 6030

RESULT 9
US-08-746-682A-1/C
; Sequence 1, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLETT, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Penile & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,682A
FILING DATE: 14-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/597,236
FILING DATE: 20-JUN-1995
APPLICATION NUMBER: EP 95201669.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci A., Allan
REGISTRATION NUMBER: 30256
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 14602 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 352..1803
```



```
RESULT 11
US-09-019-201A-1
; Sequence 1, Application US/09019201A
; Patent No. 5968780
; GENERAL INFORMATION:
; APPLICANT: FENG, PING
; APPLICANT: SOPPET, DANIEL R.
; APPLICANT: LI, YI
; APPLICANT: DILLON, PATRICK J.
; TITLE OF INVENTION: DENDRITIC CELL-DERIVED GROWTH FACTOR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,201A
; FILING DATE: HEREWITH
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: A. ANDERS BROOKES
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF346
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3621 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 134..1666
; NAME/KEY: sig_peptide
; LOCATION: 134..212
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 215..1666
; US-09-019-201A-1

Query Match 1.7%; Score 32.6; DB 2; Length 3621;
Best Local Similarity 57.3%; Pred. No. 6.2;
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 108 tattcatctaactctcttggaactttatgtgtagtctccattctctgtgcta 167
      ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1
DB 2326 TATTTCTCTCTCTCTGACACTTTCCTGGGTGCTCATTCATTCTTCTTTT 2385
      ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1

QY 168 aattccctgagtcgtgtagtcataagaatcccttatataagtaac 210
      ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1
DB 2386 CATACCAATTATTTGTTAATGATTCCACATTATTTATATGAC 2428

RESULT 12
US-08-323-443B-1/C
; Sequence 1, Application US/08323443B
; Patent No. 5654170
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W.
```

```
; APPLICANT: LANDES, GREGORY M.
; APPLICANT: BURN, TIMOTHY C.
; APPLICANT: CONNORS, TIMOTHY D.
; APPLICANT: DACKOWSKI, WILLIAM R.
; APPLICANT: GERIINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,443B
; FILING DATE: 12-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, S. Peter
; REGISTRATION NUMBER: 25,351
; REFERENCE/DOCKET NUMBER: 0372/0A462
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31571 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: PKD1 GENOMIC
; US-08-323-443B-1

Query Match 1.7%; Score 32.6; DB 1; Length 31571;
Best Local Similarity 63.3%; Pred. No. 22;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 351 gcaccatcatctggagcgcgcgtggcggaatcttcagaagccttgagcgtgtctcccaag 410
      ||| ||||| 1 ||| 1 ||| 1 ||||| 1 ||||| 1 ||||| 1
DB 5078 GCAGCATATTATMACCAAGGAGGCTGTGCTCAGAAAGCCCGCTGTCTCCACCTGG 5019
      ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1

QY 411 ctatgaactgcagacc 429
      ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1
DB 5018 AACTGGACACAGCTGTCCC 5000

RESULT 13
US-08-658-136-2/C
; Sequence 2, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERIINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
```



```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 149..9556
US-08-323-170B-1

```

```

Query Match      1.7%; Score 32.2; DB 1; Length 9636;
Best Local Similarity 49.7%; Pred. No. 15;
Matches 82; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

```

QY 16 aaaggtccatgagtggttataatgtcaaggctccccaacaaacttgaatct 75
   |||||
Db 6967 AATATGTTCACTACTTATATATCTTTATATATTTCTAATAAATTAATAACA 6908
   |||||

QY 76 tgctagagaggtlgaacatcgatcgaactgaactgaactgaactgaact 135
   |||||
Db 6907 TGCTTGTGGAATATAGCAAAACCTTTAGGACATTTTCATGCCAAATATCTCTTTT 6848
   |||||

QY 136 atgtgtagtcctcccaattcctctgtgtgtaaatlccctgagtc 180
   |||||
Db 6847 AATTTTAAAGTGCACCTTTTCTTTTACTTGACTATTTAATTC 6803
   |||||

```

Search completed: August 20, 2002, 01:09:12
Job time: 4427 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 23:00:25 ; Search time 1586.97 Seconds
(without alignments)
16218.765 Million cell updates/sec

Title: US-09-662-812-1
1907
Sequence: 1 gtgctgatttgaaag.....ctcatcaagaagcgttgaa 1907

Scoring table: IDENTITY NUC
Gapop 10.0 ; Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.8	2.5	1101	12	CNS0039G
2	42.2	2.2	1101	12	CNS017KX
3	39.6	2.1	454	9	AI742802
4	39.6	2.1	564	9	AW216328
5	39.6	2.1	579	12	AO512496
6	39.6	2.1	719	12	AZ058799
7	39.2	2.1	364	10	BF943308
8	39.2	2.1	411	10	BI343252
9	39.2	2.1	452	9	AW952059
10	39.2	2.1	454	9	AI689130
11	39.2	2.1	455	9	AW999085
12	39.2	2.1	532	9	AI684725
13	39.2	2.1	733	10	BE869217
14	38.6	2.0	339	12	AZ911277
15	38.6	2.0	487	12	AZ074909
16	38.4	2.0	466	10	BE927675
17	38.4	2.0	618	10	BF305035

C	18	38.4	2.0	724	10	BG479306	BG479306	602526335
C	19	38.2	2.0	509	12	AO695687	AO695687	HS_2159_A
C	20	38.2	2.0	685	9	AV404653	AV404653	AV404653
C	21	38.2	2.0	725	12	BH467155	BH467155	BOHOK53TR
C	22	38.2	2.0	728	12	BH466366	BH466366	BOHON53TR
C	23	38.2	2.0	1101	12	CNS016B8	AI106526	Drosophila
C	24	38.2	2.0	807	12	AZ528490	ENTCM64TR	AZ528490
C	25	38	2.0	845	12	BH139541	ENTNG88TR	BH139541
C	26	37.8	2.0	365	12	AO142345	AO142345	HS_3064_A
C	27	37.8	2.0	464	9	AV707816	AV707816	AV707816
C	28	37.6	2.0	898	9	AL574540	AL574540	AL574540
C	29	37.6	2.0	1046	12	CNS0407V	AL307444	Tetraodon
C	30	37.4	2.0	513	12	AO718260	AO718260	HS_5513_B
C	31	37.2	2.0	207	9	AI039088	AI039088	ox27601.s
C	32	37.2	2.0	275	10	D61175	D61175	HUM164D03B
C	33	37.2	2.0	647	12	BH577763	BH577763	BOGGA777R
C	34	37.2	2.0	725	12	AZ200838	AZ200838	SP_1011_A
C	35	37	1.9	533	12	AQ420915	AQ420915	RPCT-11-2
C	36	37	1.9	851	10	BF132845	BF132845	601645986
C	37	36.8	1.9	527	9	AJ398606	AJ398606	AJ398606
C	38	36.8	1.9	730	12	AZ519102	AZ519102	RPCT-11-7
C	39	36.8	1.9	900	9	AU067211	AU067211	AU067211
C	40	36.6	1.9	479	12	AQ002804	AQ002804	CIT-HSP-2
C	41	36.6	1.9	542	10	BI935367	BI935367	EST55256
C	42	36.6	1.9	740	10	BI926005	BI926005	EST545894
C	43	36.6	1.9	794	10	BI935368	BI935368	EST55257
C	44	36.4	1.9	421	9	AI267708	AI267708	ag93109.x
C	45	36.4	1.9	461	9	AW365573	AW365573	MR0-HR007

ALIGNMENTS

RESULT 1
CNS0039G
LOCUS
DEFINITION
CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL063921
VERSION
AL063921.1 GI:4941778
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCT-98"

QY	134	ctatgctgagaggttgcgaacatcgaatcgaacagcgagccttagcaacatcctggagc	249
DB	272	GCATTGTAAAGAGTACACTCTGTCTTCAAAACGTATCTTCTTGGATGGAATTAAGATG	331
QY	194	atcccttataatgaatgcgaatcgaacagcgagccttagcaacatcctggagc	249
DB	332	TAACTGTATAGTTTAAAGATTAATTAATGGAAGTTGGTCCTCAACTAAGATGACAGC	387
RESULT	11		
AM999085			
LOCUS	AM999085	455 bp	mRNA
DEFINITION	MR0-BN0070-010300-003-e11 BN0070 Homo sapiens CDNA, mRNA sequence.		linear
ACCESSION	AM999085		
VERSION	AM999085.1	GI:8259319	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
	1 (bases 1 to 455)		
	Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R.,		
	Nagai, M. A., da Silva, W. J., Zago, M. A., Bordin, S., Costa, F. F.,		
	Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,		
	Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare		
	, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and		
	Simpson, A. J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed		
	sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
MEDLINE	20202663		
COMMENT	Contact: Simpson A.J.G.		
	Laboratory of Cancer Genetics		
	Ludwig Institute for Cancer Research		
	Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,		
	Brazil		
	Tel: +55-11-2704922		
	Fax: +55-11-2707001		
	Email: asimpson@ludwig.org.br		
	This sequence was derived from the FAPESP/LICR Human Cancer Genome		
	Project. This entry can be seen in the following URL		
	(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=st2-MR0-BN0070-010300-003-e11&t3=2000-03-01&t4=1)		
	Seq primer: puc 18 forward		
	High quality sequence start: 13		
	High quality sequence stop: 455.		
FEATURES			
source	Location/Qualifiers		
	1..455		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="BN0070"		
	/dev_stage="Adult"		
	/note="Organ: breast,normal; Vector: puc18; Site:1: SmaI;		
	Site:2: SmaI; A mini-library was made by cloning products		
	derived from ORESTES PCR (U.S. Letters Patent application		
	No. 196,716 - Ludwig Institute for Cancer Research)		
	profiles into the puc 18 vector. Reverse transcription of		
	tissue mRNA and cDNA amplification were performed under		
	low stringency conditions."		
BASE COUNT	163 a 75 c 64 g 153 t		
ORIGIN			
	Query Match 2.1%; Score 39.2; DB 9; Length 455;		
	Best Local Similarity 47.9%; Pred. No. 5.9;		
	Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;		
QY	14	gaaaagagtcgaatgagtggtttataatggttcgaagtcctccctcaacaaacttgaata	73
DB	192	GAATAAATAATATACCTTAAACATTCAGTTAAATATGTAAGTTTATCTAAATCTCAAAATG	251
QY	74	cttgctagagaggttgcgaacatcgaatcgaacagcgagccttagcaacatcctggagc	249
DB	332	TAACTGTATAGTTTAAAGATTAATTAATGGAAGTTGGTCCTCAACTAAGATGACAGC	387

Db	252	TTTAAATAAAACAGTACTCTTCOCATTTTAACACTCTTCTTCTACAGTACAGTAAT	311
Oy	134	ttattgtgtagctcttcccatcttctctgtcttaattccctagctgtagtcataag	193
Db	312	GCATTTAGAGAGTACACTTCTGCTTCAACCTGATCTTTTGAGATGGAATTAAGATG	371
Oy	194	atcccttataatagtaacgaactcacacagcagcgcccttagcacacatctcgaagc	249
Db	372	TAACTGTACTTTTAAGATTAATTAATGGAAGTGTGTCACACTAAGATGACAGC	427
RESULT	12		
AI684725			
DEFINITION	AI684725	532 bp	mRNA linear EST 07-MAR-2000
ACCESSION	U855409.x1	Soares,NFL,T.GRC.S1	Homo sapiens cDNA clone
VERSION	AI684725	IMAGE:2302961.3	, mRNA sequence.
KEYWORDS	AI684725.1	GI:4896019	
ORGANISM			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

FEATURES		SOURCE	
		Location/Qualifiers	
		1..532	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:2302961"	
		/clone_1lb="Soares_NFL_T_GBC_S1"	
		/lab_host="DH10B"	
		/note="Organ: pooled; Vector: pF73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBL19w, testis NHT, and B-cell NCI-GAP.GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pool consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonalogo."	
BASE COUNT		198 a	82 c 74 g 178 t
ORIGIN			
Query Match	2.1%; Score 39.2; DB 9; Length 532;		
Best Local Similarity	47.9%; Pred. No. 6.2;		
Matches 113; Conservative	0; Mismatches 123; Indels 0; Gaps 0;		
OY	14	gaaaaagtcacatgcgatgtgtttataatgltccaagtcctccatccaacaacttgaata	73
Db	147	gaaaaaagataatcatcctttagacattcagtttataatgtatgtttatcttaattctcaaaatg	206
OY	74	ctctgtagaggaagtgaacatgcatactggaacatccatccaactccttgcact	133
Db	207	tttataataaaacaactatcttctccattttaacaccttttgcctttcttactgctacagtaatt	266
OY	134	ttattgttagtcctccacttctctgtgttgcttaaatccctcgaagctctgtagtcataag	193
Db	267	gcattgtagagagctacacttctcttcaaaactgtatctcttttgatggaaatttaagatg	326

Db	327	TAACGTATAGTTTAAAGATAAATAAATGGGAGATTGGTCCACTAAGATGACACG	382
RESULT	13		
LOCUS	BE869217/c		
DEFINITION	BE869217	733 bp	mRNA linear EST 20-OCT-2000
ACCESSION	6014454023p1 NIH_MGC_65	Homo sapiens	CDNA clone IMAGE:3849388 5',
VERSION	BE869217		
KEYWORDS	BE869217..1	GI:10317993	
SOURCE	EST.		
ORGANISM	human.		
	Homo sapiens		
	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 733)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLM9566 row: m column: 05 High quality sequence stop: 698.		

FEATURES	source	location/Qualifiers
		1..733
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="IMAGE:3849388"
		/clone_1ib="NH.MGC.65"
		/tissue.type="adenocarcinoma"
		/lab_host="DH10B (phage-resistant)"
		/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
		Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
		Average insert size 1.8 kb. Library constructed by Life
		Technologies."
BASE COUNT	237 a	110 c 136 g 249 t 1 others
ORIGIN		
Query Match	2.1%;	Score 39.2; DB 10; Length 733;
Best Local Similarity	47.9%;	Pred.No. 7;
Matches 113;	Conservative 0;	Mismatches 123; Indels 0; Gaps 0;
QY	14	gaaaaagtcctcgtgatgtgttataaagttcaagtcctccctaccaacatgaata 73
DB	257	GAAAAAGGTAATTCACCTTACGACATTCAGTTAAATGTAAGTTTATCTCAATAATG 198
QY	74	cttgctagagaggttgaacatcgatcctatggaagattccactcactcctcttggact 133
DB	197	TTTAAATAAAAACAAGATATCTTCCATCTTAAACACCTTGCTTCTCAAGCTACAGTAAT 138
QY	134	ttatgtgtagtctctccactctctctgtgtcttaaatccctcgagtcgtltaggtcataag 193
DB	137	GCATTGTAGAGAGTACACTTCGTGCTTCAAACTGTATCTCTTTGGATGGAAFTTAAGATG 78
QY	194	atcccttatataagtaagcaatcctacacagcagcgcttagacaacatcttgaagc 249
DB	77	TACCTGTATAGTTTAAAGTAAATTAATGGAAGTTGGCTCAACATGAATGACACG 22
RESULT 14		
AC2911277/c	A2911277	339 bp DNA 1linear GSS 05-MAR-2001

DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
RPCT-24-222010.TJ RPCT-24	RPCT-24-222010	A2911277	GSS.	house mouse.	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akincet, B., Levins, M., Tseng, G., Geer, K., Krol, M., Shartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C. M.	Mouse BAC End Sequences from Library RPCT-24 unpublished (1998)	Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhaoc@ig.org	
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						
Best Local Similarity	58.1%	Pred. No. 7.9.	Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;						
Query Match	2.0%	Score 38.6;	DB 12; Length 339;						

```

REFERENCE      1 (bases 1 to 487)
AUTHORS        Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akintet
                ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroll,M., de Jong,P.
                and Fraser,C.M.
TITLE          Mouse BAC End Sequences from Library RPCI-23
JOURNAL        Unpublished (1999)
COMMENT        Other-SSS: RPCI-23-438022.TV
                Contact: Shaying Zhao
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: szhao@tigr.org
                Clones are derived from the mouse BAC library RPCI-23. For BAC
                library availability, please contact Pieter de Jong
                (pieter@dejong.med.buffalo.edu). Clones may be purchased from
                BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
                or from Resea ch Genetics (Info@resgen.com). BAC end page:
                http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
                Plate: 438 row: 0 column: 22
                Seq primer: SP6
                Class: BAC ends.

FEATURES
Source         location/qualifiers
               1..487
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="RPCI-23-438022"
               /clone_lib="RPCI-23"
               /sex="female"
               /lab_host="DH10B"
               /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1:
               EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
               brain genomic DNA was isolated and partially digested
               with a combination of EcoRI and EcoRI Methylase. Size
               selected DNA was cloned into the pBACE3.6 vector at the
               EcoRI sites. The ligation products were transformed into
               DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT    105 a      100 c      140 g      142 t
ORIGIN
Query Match   2.0%; Score 38.6; DB 12; Length 487;
Best Local Similarity 58.1%; Pred.No. 9;
Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
OY 649 acgccttgagagctgaagatctgtagcttatattccggagatctcttagctgccagaagaag 708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 ACACGGAAGACTGGAAGAACTGCTTCGGAATCAGGAGGAACGAAACACACACACAG 293
OY 709 cgcgattcgaggtgccacagcttgcagatcgagagaatcaccacaagaagcttct 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 GGCAGATGAGCTGCGCAGACGCAATGTGAAGGTAAACGACCCAAAACCAAGCTCTCT 236

Search completed: August 20, 2002, 00:26:11
Job time: 5146 sec

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:22:21 ; Search time 19.49 seconds
(without alignments)
2800.346 Million cell updates/sec

Title: US-09-662-812-2

Perfect score: 568

Sequence: 1 MGLFHLTFGLLCSLPISL.....TEGDAKNFPVLAKLTIKIVE 568

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: pirl:.*
2: pirl2:.*
3: pirl3:.*
4: pirl4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568	100.0	568	2 C72129	probable outer mem
2	568	100.0	568	2 G86493	probable leader pe
3	12	2.1	566	2 G71525	probable outer mem
4	10	1.8	566	2 A81682	conserved hypochet
5	9	1.6	408	1 E42409	biphenyl dioxygena
6	9	1.6	408	1 F41858	probable glucose-i
7	9	1.6	445	2 T03578	hypothetical prote
8	9	1.6	502	2 T05135	hypothetical prote
9	8	1.4	134	2 S77785	hypothetical prote
10	8	1.4	137	2 C81695	cyclic nucleotide-
11	8	1.4	197	2 E95203	conserved hypochet
12	8	1.4	197	2 E98070	conserved hypochet
13	8	1.4	308	2 A12601	conserved hypochet
14	8	1.4	328	2 A97384	hypothetical prote
15	8	1.4	357	2 T03557	ribose transport s
16	8	1.4	414	2 S52618	hypothetical prote
17	8	1.4	416	2 G71096	hypothetical prote
18	8	1.4	434	2 A81597	similar glucose in
19	8	1.4	434	2 AD1234	glucose-inhibited
20	8	1.4	435	2 A72339	glucose-inhibited
21	8	1.4	435	2 A69632	glucose-inhibited
22	8	1.4	435	2 B83958	glucose-inhibited
23	8	1.4	435	2 F89898	glucose-inhibited
24	8	1.4	444	2 D97977	glucose-inhibited
25	8	1.4	447	2 A64934	Succinylarginine d
26	8	1.4	447	2 C90935	hypothetical prote
27	8	1.4	447	2 G85783	hypothetical prote
28	8	1.4	447	2 G86778	glucose inhibited
29	8	1.4	456	2 A95109	Gid protein [impor

30	8	1.4	483	2 T21327	hypothetical prote
31	8	1.4	566	2 T45162	succinate dehydrog
32	8	1.4	587	2 C86744	myosin-crossreacti
33	8	1.4	588	2 S66625	zeta-carotene desa
34	8	1.4	629	2 D96533	ARP protein [impor
35	8	1.4	629	2 S57614	ARP protein - Arab
36	8	1.4	666	2 F71032	flagellar hook-ass
37	8	1.4	667	2 F70682	probable membrane
38	8	1.4	671	2 A10816	conserved hypochet
39	8	1.4	761	2 H84950	ribonucleoside-dip
40	8	1.4	998	2 T32787	hypothetical prote
41	7	1.2	54	2 S60852	M protein precuso
42	7	1.2	76	2 E58933	ATP synthase prote
43	7	1.2	80	4 S60901	hypothetical prote
44	7	1.2	103	2 D72521	hypothetical prote
45	7	1.2	104	2 D72657	hypothetical prote

ALIGNMENTS

RESULT 1

probable outer membrane leader peptide (comp) CPn0021 - Chlamydomophila pneumoniae (str C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C/Accession: C72129, B81542

R/Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999

M/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A/Reference number: A72000; MUID:99206606

A/Accession: C72129

A/Molecule type: DNA

A/Residues: 1-568 <ARN>

A/Cross-references: GB:AE001587; GB:AE001363; NID:94376271; PIDN:AD18174.1; PID:943

A/Experimental source: strain CWL029

R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hick Nucleic Acids Res. 28, 1397-1406, 2000

M/Title: Genome sequences of Chlamydia trachomatis Mofn and Chlamydia pneumoniae AR3

A/Reference number: A81500; MUID:20150255

A/Accession: B81542

A/Molecule type: DNA

A/Residues: 1-568 <RNA>

A/Cross-references: GB:AE002234; GB:AE002161; NID:97189667; PIDN:AR38558.1; PID:971

A/Experimental source: strain AR39, HL cells

C/Genetics:

A/Gene: CPn0021; CP0755

Query Match	Score	DB	Length	568:
Best Local Similarity	100.0%	Pred. NO. 0;		
Matches	568:	Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
1	MGLFHLTFGLLCSLPISLVAKFPESYGHILYISTOSTOQALATYLEALDAYGDHDF	60		
1	MGLFHLTFGLLCSLPISLVAKFPESYGHILYISTOSTOQALATYLEALDAYGDHDF	60		
61	VLRKIGEDYLKQSHSSDPQTRKSTTIGAGLGSSEALDVLVSQMETADPLDQLLVLSAV	120		
61	VLRKIGEDYLKQSHSSDPQTRKSTTIGAGLGSSEALDVLVSQMETADPLDQLLVLSAV	120		
61	VLRKIGEDYLKQSHSSDPQTRKSTTIGAGLGSSEALDVLVSQMETADPLDQLLVLSAV	120		
121	SGHIGKTSDDLFRALASPVYIRLEAAYRIANLKNFVVIDHLHSFHKLPEDIOCLSA	180		
121	SGHIGKTSDDLFRALASPVYIRLEAAYRIANLKNFVVIDHLHSFHKLPEDIOCLSA	180		
121	SGHIGKTSDDLFRALASPVYIRLEAAYRIANLKNFVVIDHLHSFHKLPEDIOCLSA	180		
181	IFLRLEFEESDAYIRDLAAKSAIRSATATQIGGYOQKRELPTLRNLTLTSASPODOBAI	240		
181	IFLRLEFEESDAYIRDLAAKSAIRSATATQIGGYOQKRELPTLRNLTLTSASPODOBAI	240		
181	IFLRLEFEESDAYIRDLAAKSAIRSATATQIGGYOQKRELPTLRNLTLTSASPODOBAI	240		
241	LYALGKTLKDGOSYNIKKQLOKPDVDYTLAAQALIALGKEDALPVYIKKQALERRPAL	300		
241	LYALGKTLKDGOSYNIKKQLOKPDVDYTLAAQALIALGKEDALPVYIKKQALERRPAL	300		
241	LYALGKTLKDGOSYNIKKQLOKPDVDYTLAAQALIALGKEDALPVYIKKQALERRPAL	300		

Qy	301	YAHHPBSEICITPALPPIFLTKNKSSEKALNVALALLEGCDTRKLEYITERLVORHANE	360
Db	301	YALRLPBEICIPALPPIFLTKNKSSEKALNVALALLEGCDTPKLEYITERLVORPHANE	360
Qy	361	TLASFSGKRTLOMKRNIIIVPODDPQERRELTSTRGLEQILTEFLRPLKCAVLPCTI	420
Db	361	TLASFSGKRTLOMKRNIIIVPODDPQERRELTSTRGLEQILTEFLRPLKCAVLPCTI	420
Qy	421	KLASQRTQLATTAISFISHSHOEADLDLFOAKLPGEPITIRAYADLAIYNLTKDPEKK	480
Db	421	KLASQRTQLATTAISFISHSHOEADLDLFOAKLPGEPITIRAYADLAIYNLTKDPEKK	480
Qy	481	RLSHDYAKKLIQOEILFEDFENQRPBMSPLRQVTPSERTKLMJDILETLATSKSED	540
Db	481	RLSHDYAKKLIQOEILFEDFENQRPBMSPLRQVTPSERTKLMJDILETLATSKSED	540
Qy	541	IRLLIQMTGEGDAKNFPVLAGLLIKIYE	568
Db	541	IRLLIQMTGEGDAKNFPVLAGLLIKIYE	568

RESULT 2

probable leader peptide Omp [imported] - Chlamydoiphila pneumoniae (strain J138)
C:Species: Chlamydoiphila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: G86493
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Iwano
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: G86493
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1,568 <STD>
A:Cross-references: GB:BA000008; NID:98978396; PIDN:BA9823.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
C:Gene: CPJ0021

Query Match	100.0%	Score 568:	DB 2:	Length 568:
Best Local Similarity	100.0%	Pred. No. 0:		
Matches 568:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:
QY 1	MGLFHLTFLGGLLCSLP	SLVAKPEPESVGHKILY	ISTOSTOOLATY	EALDAYGDHDF 60
Db 1	MGLFHLTFLGGLLCSLP	SLVAKPEPESVGHKILY	ISTOSTOOLATY	EALDAYGDHDF 60
QY 61	VLKRIGEDYLYKQSH	HSDDPQTRKSTI	IGAGLAGSSEALDVL	SQAMETADPLQQLLVLSAV 120
b 61	VLKRIGEDYLYKQSH	HSDDPQTRKSTI	IGAGLAGSSEALDVL	SQAMETADPLQQLLVLSAV 120
QY 121	SGHLGKTSDDLFLFK	LAPPYPIYIREAAVYRL	ANKNTYVIOHLHSF	HKLPREIOCLSSAA 180
Db 121	SGHLGKTSDDLFLFK	LAPPYPIYIREAAVYRL	ANKNTYVIOHLHSF	HKLPREIOCLSSAA 180
QY 181	IFLRLTEESDAYIR	LDLLAAKSAIRSAT	ALQIGEYQOKRELP	PLTNLLTSASPODOEAI 240
Db 181	IFLRLTEESDAYIR	LDLLAAKSAIRSAT	ALQIGEYQOKRELP	PLTNLLTSASPODOEAI 240
QY 241	LYALGKLTDDGQSY	NIKKOLQKPDVDVY	LAAQAALIAIGKED	DALPVYIKKQALBERPRAL 300
Db 241	LYALGKLTDDGQSY	NIKKOLQKPDVDVY	LAAQAALIAIGKED	DALPVYIKKQALBERPRAL 300
QY 301	YALRHLPEIEIGIP	IALPIFLFKTKNSEAK	LANVALLEIGCDTP	KKLEXYITERLYVPHNE 360
Db 301	YALRHLPEIEIGIP	IALPIFLFKTKNSEAK	LANVALLEIGCDTP	KKLEXYITERLYVPHNE 360
QY 361	TLALSFSGRFLQNM	KRNIIIVPODPQER	RLSTTGLBEOLIT	FLFRLPREKVALPCIT 420
Db 361	TLALSFSGRFLQNM	KRNIIIVPODPQER	RLSTTGLBEOLIT	FLFRLPREKVALPCIT 420

QY	421	KLASQKQOLATTAISFSSHTSHOEADLLFOAKLPEEPPIRAYADLATINLTKDEPK	480
Db	421	KLASQKQOLATTAISFSSHTSHOEADLLFOAKLPEEPPIRAYADLATINLTKDEPK	480
QY	481	RSJLHYAKKLIQETLLFQDTENQRPHPRLRYQVPESTKIMLDILETLATSKSED	540
Db	481	RSJLHYAKKLIQETLLFQDTENQRPHPRLRYQVPESTKIMLDILETLATSKSED	540
QY	541	IRLLIOLMTESGAKNFPVLAGLLIKIVE	568
Db	541	IRLLIOLMTESGAKNFPVLAGLLIKIVE	568

RESULT 3

probable outer membrane leader peptide (omp) CT350 - Chlamydia trachomatis (serotype
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence-revision 13-Sep-1998 #text-change 04-Feb-2000
 C:Accession: G71525
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
 Science 282, 754-759, 1998
 A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
 A:Reference number: A71570; MUID:99000809
 A:Accession: G71525
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-566 <ARN>
 A:Cross-references: GB:AE001308; GB:AE001273; NID:g3328766; PIDN:AMC67945.1; PID:g332
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: CT350

Query Match	2.1%	Score 12	DB 2	Length 566
Best Local Similarity	100.0%	Pred. No.	0.0033	
Matches 12	Conservative	0	Mismatches 0	Indels 0
0y	141	PVIRLEAAVRLA	152	
Db	139	PVIRLEAAVRLA	150	

RESULT 4

conserved hypothetical protein TC0629 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: A81682
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: A81682
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-566 <TEF>
A:Cross-references: GB:AE002331; GB:AE002160; NID:g71900662; PIDN:AAF39458.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0629

	Query Match	1.8%;	Score 10;	DB 2;	Length 566;
	Best Local Similarity	100.0%;	Pred. No. 0.32;		
	Matches 10;	Conservative	0;	Mismatches	0;
				Indels	0;
					Gaps 0;
QY	143	IRLEAAAYRLA	152		
DB	141	IRLEAAAYRLA	150		

RESULT 5
E42409
biphenyl dioxygenase (EC 1.14.-.-) ferredoxin reductase component - Pseudomonas pseudoal
C:Species: Pseudomonas pseudoalcaligenes
C>Date: 04-Mar-1993 #sequence_revision 06-Jan-1995 #text_change 11-Jun-1999
C:Accession: E42409
R:Taiba, K.; Hirose, J.; Hayashida, S.; Furukawa, K.
J. Biol. Chem. 267, 4844-4853, 1992
A:Title: Analysis of bph operon from the polychlorinated biphenyl-degrading strain of P
A:Reference number: A42409; MUID:92165849
A:Accession: E42409
A:Molecule type: DNA
A:Residues: 1-408 <TAI>
A:Cross-references: GB:M83673; NID:g151090; PIDN:AA25747.1; PID:g151095
A:Experimental source: strain KF707
A>Note: sequence extracted from NCBI backbone (NCBIN:84014, NCBIP:84021)
C:Genetics:
A:Gene: bphA4; bphG
C:Superfamily: toluene dioxygenase ferredoxin reductase component
C:Keywords: aromatic hydrocarbon catabolism; FAD; NAD; oxidoreductase; PCB biodegradation
F:2-408/Product: biphenyl dioxygenase ferredoxin reductase component #status predicted <
F:9-25/Region: FAD/NAD-binding motif
F:150-166/Region: FAD/NAD-binding motif
F:265-278/Region: FAD/NAD-binding motif

Query Match 1.6%; Score 9; DB 1; Length 408;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 ITAGLAGS 94
|||||
DB 7 ITAGLAGS 15

RESULT 6
F41858
biphenyl dioxygenase (EC 1.14.-.-) ferredoxin reductase component - Pseudomonas sp.
C:Species: Pseudomonas sp.
C>Date: 04-Mar-1993 #sequence_revision 06-Jan-1995 #text_change 10-Feb-1995
C:Accession: F41858
R:Ericksen, B.D.; Mondello, F.J.
J. Bacteriol. 174, 2903-2912, 1992
A:Title: Nucleotide sequencing and transcriptional mapping of the genes encoding biphenyl
A:Reference number: A41858; MUID:92234948
A:Accession: F41858
A:Molecule type: DNA
A:Residues: 1-408 <ERI>
A:Experimental source: strain LB400
A>Note: sequence extracted from NCBI backbone (NCBIN:97256, NCBIP:97265)
C:Genetics:
A:Gene: bphA4; bphG
C:Superfamily: toluene dioxygenase ferredoxin reductase component
C:Keywords: aromatic hydrocarbon catabolism; FAD; NAD; oxidoreductase; PCB biodegradation
F:2-408/Product: biphenyl dioxygenase ferredoxin reductase component #status predicted <
F:9-25/Region: FAD/NAD-binding motif
F:150-166/Region: FAD/NAD-binding motif
F:265-278/Region: FAD/NAD-binding motif

Query Match 1.6%; Score 9; DB 1; Length 408;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 ITAGLAGS 94
|||||
DB 7 ITAGLAGS 15

RESULT 7
T03578
probable glucose-inhibited division protein gid - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus

C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03578
R:Ylcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SBI
A:Reference number: Z14955; MUID:97404404
A:Accession: T03578
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1445 <YLC>
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AC16231.1; PID:g3128379
C:Genetics:
A:Map position: 1

Query Match 1.6%; Score 9; DB 2; Length 445;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 ITAGLAGS 94
|||||
DB 6 ITAGLAGS 14

RESULT 8
T05135
hypothetical protein F7H19.220 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05135
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.;
submitted to the Protein Sequence Database, July 1998
A:Reference number: Z15399
A:Accession: T05135
A:Molecule type: DNA
A:Residues: 1-502 <BEV>
A:Cross-references: EMBL:AL031018
A:Experimental source: cultivar Columbia; BAC clone F7H19
C:Genetics:
A:Map position: 4
A>Note: F7H19.220

Query Match 1.6%; Score 9; DB 2; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 LLICSLPTS 19
|||||
DB 122 LLICSLPTS 130

RESULT 9
S77785
hypothetical protein MC127 - Mycoplasma capricolum
C:Species: Mycoplasma capricolum
C>Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 07-Dec-1999
C:Accession: S77785
R:Bork, P.; Ouzounis, C.; Casari, G.; Schneider, R.; Sander, C.; Dolan, M.; Gilbert, M.;
Mol. Microbiol. 16, 955-967, 1995
A:Title: Exploring the Mycoplasma capricolum genome: a minimal cell reveals its phys
A:Reference number: S77739; MUID:96059641
A:Accession: S77785
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-134 <BOR>
A:Cross-references: EMBL:Z33089; NID:g514453; PIDN:CA83753.1; PID:g530429
A:Experimental source: ATCC 27343
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Genetics:
A:Genetic code: SGC3

Query Match 1.4%; Score 8; DB 2; Length 134;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 86 IIGAGLAG 93
 Db 7 IIGAGLAG 14

RESULT 10
 C81695
 cyclic nucleotide-binding protein, probable TC0506 [imported] - Chlamydia muridarum (str
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: C81695
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255
 A:Accession: C81695
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-137 <TE>
 A:Cross-references: GB:AE002319; GB:AE002160; NID:g7190541; PIDN:AAE39348.1; PID:g719054
 A:Experimental source: strain N19g (Mopn)
 C:Genetics:
 A:Gene: TC0506

Query Match 1.4%; Score 8; DB 2; Length 137;
 Best Local Similarity 100.0%; Pred. No. 8.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 331 VALALLEL 338
 Db 120 VALALLEL 127

RESULT 11
 E95203
 conserved hypothetical protein SPI746 [imported] - Streptococcus pneumoniae (strain TIGC
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: E95203
 R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hel
 nson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfe,
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: E95203
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-197 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK75822.1; PID:g14973242; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SPI746
 C:Superfamily: hypothetical protein ygek

Query Match 1.4%; Score 8; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 LHDYAKKL 490
 Db 57 LHDYAKKL 64
 RESULT 12

E98070
 conserved hypothetical protein spi1591 [imported] - Streptococcus pneumoniae (strain
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: E98070
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: E98070
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-197 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAU00394.1; PID:g15459258; GSPDB:GN00174
 C:Genetics:
 A:Gene: spi1591
 C:Superfamily: hypothetical protein ygek

Query Match 1.4%; Score 8; DB 2; Length 197;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 483 LHDYAKKL 490
 Db 57 LHDYAKKL 64

RESULT 13
 A12601
 conserved hypothetical protein Atu0208 [imported] - Agrobacterium tumefaciens (strain
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: A12601
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutlyavin, T.; Levy, R.; Li, M.; McCl
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: A12601
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-308 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAU41231.1; PID:g17738535; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu0208
 A:Map position: circular chromosome

Query Match 1.4%; Score 8; DB 2; Length 308;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAG 93
 Db 189 IIGAGLAG 196

RESULT 14
 A97384
 hypothetical protein AGR_C357 [imported] - Agrobacterium tumefaciens (strain C58, Ce
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: A97384
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldm
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Leppas, C.; Markelz,

Science 294, 2323-2328, 2001
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*
 A:Reference number: A97359; PMID:11743194
 A:Accession: A97384
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-328 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK86026.1; PID:g15155095; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR.C.357
 A:Map position: circular chromosome

Query Match 1.4%; Score 8; DB 2; Length 328;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TTGAGTAG 93
 |||||
 Db 209 TTGAGTAG 216

RESULT 15
 T03557
 ribose transport system permease protein RbSC - *Rhodobacter capsulatus*
 C:Species: *Rhodobacter capsulatus*
 C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Nov-1999
 C:Accession: T03557
 R:Vleck, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
 Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
 A>Title: Sequence of a 189-kb segment of the chromosome of *Rhodobacter capsulatus* SB1003
 A:Reference number: 214955; MUID:97404404
 A:Accession: T03557
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-357 <VLC>
 A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAQ16210.1; PID:g3128358
 C:Genetics:
 A:Map position: 1
 C:Superfamily: 1-arabinose transport system permease arah
 C:Keywords: ribose transport

Query Match 1.4%; Score 8; DB 2; Length 357;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 LAAAGALI 276
 |||||
 Db 86 LAAAGALI 93

Search completed: August 20, 2002, 04:26:37
 Job time: 256 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:26:01 ; Search time 17.08 Seconds

(Without alignments)
1287.629 Million cell updates/sec

Title: US-09-662-812-2

Sequence: 1 MGLFHLTFGLLCSLPISL.....TEGAKNPVLAGLLIKIVE 568

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.6	408	1	BPHG_BURCE
2	8	1.4	342	1	SNXF_HUMAN
3	8	1.4	435	1	GID_BACSU
4	8	1.4	447	1	ASTB_ECOLI
5	8	1.4	558	1	ZDS_ARATH
6	8	1.4	570	1	ZDS_MAZE
7	8	1.4	574	1	ZDS_MARPS
8	8	1.4	587	1	ZDS_TAREP
9	8	1.4	588	1	ZDS_CAPAN
10	8	1.4	588	1	ZDS_LYCES
11	8	1.4	666	1	FLDP_VIBPA
12	8	1.4	761	1	RIRI_BUCAI
13	8	1.2	85	1	GON2_DICLA
14	7	1.2	85	1	GON2_MORSA
15	7	1.2	106	1	VMEH_CVB
16	7	1.2	122	1	SAA1_RABIT
17	7	1.2	122	1	SAA2_RABIT
18	7	1.2	122	1	SSAM_SALTY
19	7	1.2	137	1	RT16_HUMAN
20	7	1.2	147	1	PA24_BUNMU
21	7	1.2	148	1	YEES_ECOLI
22	7	1.2	163	1	LEUD_PYRHO
23	7	1.2	164	1	RECK_TREPA
24	7	1.2	186	1	YJ89_MYCTU
25	7	1.2	187	1	NFX8_PSEAE
26	7	1.2	207	1	GL21_ARATH
27	7	1.2	213	1	KGUA_CAVCR
28	7	1.2	230	1	FLAH_METVO
29	7	1.2	259	1	YCBC_ECOLI
30	7	1.2	260	1	ATP6_BRANA
31	7	1.2	286	1	PUR7_PASMU
32	7	1.2	291	1	ATP6_MAZE
33	7	1.2	306	1	PUR7_HAEIN

34	7	1.2	306	1	YDS4_SCHPO	014180 schizosach
35	7	1.2	309	1	HCD2_CAEEL	P41938 caenorhabd1
36	7	1.2	321	1	EUM1_EUMMA	P25780 eurog1yphus
37	7	1.2	324	1	PHT2_PSEPU	005182 pseudomonas
38	7	1.2	329	1	DONS_MOUSE	09qxp4 mus musculu
39	7	1.2	331	1	IPNS_PENCH	P08703 penicillium
40	7	1.2	332	1	Y675_TREPA	083681 treponema p
41	7	1.2	371	1	CYB_ASPEE	048017 aspidites m
42	7	1.2	371	1	CYB_LIALA	048090 liasis albe
43	7	1.2	371	1	CYB_LIACH	048092 liasis albe
44	7	1.2	371	1	CYB_LIAMA	048093 liasis macu
45	7	1.2	371	1	CYB_LIAMS	048094 liasis mack

ALIGNMENTS

RESULT 1
BPHG_BURCE STANDARD; PRT; 408 AA.
ID BPHG_BURCE
AC P37337;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Biphenyl dioxygenase system ferredoxin--NAD(+) reductase component
(EC 1.18.1.3).
GN BPHG.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_Taxid:292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LE400;
RX MEDLINE=92234948; Pubmed=1569021;
RA Erickson B.D., Mondello F.J.;
RT "Nucleotide sequencing and transcriptional mapping of the genes
RT encoding biphenyl dioxygenase, a multicomponent
RT polychlorinated biphenyl-degrading enzyme in Pseudomonas strain
RT LE400".
RT J. Bacteriol. 174:2903-2912(1992).
RN [2]
RP REVISIONS.
RA Erickson B.D., Mondello F.J.
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: PART OF THE ELECTRON TRANSFER COMPONENT OF BIPHENYL
DIOXYGENASE. TRANSFERS ELECTRONS FROM FERREDOXIN (BPHF) TO
NADH.
CC CATALYTIC ACTIVITY: Reduced ferredoxin + NAD(+) = oxidized
ferredoxin + NADH.
CC COFACTOR: FAD.
CC -!- PATHWAY: BIPHENYL-POLYCHLORINATED BIPHENYL DEGRADATION PATHWAY.
CC -!- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE
TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (BPHA AND BPHF), A
FERREDOXIN (BPHF) AND A FERREDOXIN REDUCTASE (BPHG).
CC -!- SIMILARITY: TO OTHER BACTERIAL RING-HYDROXYLATING DIOXYGENASE
FERREDOXIN REDUCTASE COMPONENTS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M86348; AAB63429.1; -.
CC PIR: F41858.
CC InterPro: IPR001327; FAD_Pyr_redox.
CC InterPro: IPR000205; NAD_Pyr_redox.
CC PIR: PF00070; Pyr_redox; 1.
CC PRINTS: PR00368; FADPNR.
CC Aromatic hydrocarbons catabolism; Flavoprotein; FAD; NAD;

KW Oxidoreductase. 35 FAD (ADP PART) (POTENTIAL).
 FT NP_BIND 145 173 NAD (ADP PART) (POTENTIAL).
 SO SEQUENCE 408 AA; 42953 MW; 8A52BB01688667A9 CRC64;

Query Match 1.6%; Score 9; DB 1; Length 408;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGIAGS 94
 DB 7 IIGAGIAGS 15

RESULT 2
 SNXF_HUMAN

ID SNXF_HUMAN STANDARD; PRT; 342 AA.

AC O9NRS6; O9NRS5;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sorting nexin 15.

GN SNX15.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RA [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Parathyroid;

RE MEDLINE=21265032; PubMed=11085978;

RA Phillips S.A., Barr V.A., Haft D.H., Taylor S.I., Haft C.R.;

"Identification and characterization of snx15, a novel sorting nexin

involved in protein trafficking".

J. Biol. Chem. 276:5074-5084(2001).

CC -1- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR

TRAFFICKING. OVEREXPRESSION OF SNX15 DISRUPTS THE NORMAL

TRAFFICKING OF PROTEINS FROM THE PLASMA MEMBRANE TO RECYCLING

ENDOSOMES OR THE TGN.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2/SNX15A;

CC ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.

CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.

CC -1- SIMILARITY: CONTAINS 1 PHOX HOMOLOG (PX) DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

CC -----

DR EMBL: AF175267; AAF89955.1; -

DR EMBL: AF175268; AAF89956.1; -

DR MIM: 605964; -

DR InterPro: IPR001683; PX.

DR Pfam: PF00787; PX; 1.

DR SMART: SM00312; PX; 1.

KW Transport; Protein transport; Alternative splicing.

FT DOMAIN 11 123 PX.

FT VARSPLIT 221 308

FT EEGAAPSPTVAALATNEVASRLQDEPWPFGQEEEDGE

FT GGPVPAVLSQATELITQALNDEKAGAYAAALOGYRDGVHYL

FT LOGVPS -> G (IN ISOFORM 2).

SEQUENCE 342 AA; 38291 MW; 33F64A79EAF6BDC CRC64;

OY 444 QEADLDF 451

Query Match 1.4%; Score 8; DB 1; Length 342;

Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|||||||
 143 QEADLDF 140

RESULT 3

GID_BACSU

ID GID_BACSU STANDARD; PRT; 435 AA.

AC P39815;

DT 01-FEB-1995 (Rel. 31, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE protein gid.

GN GID.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RA [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Foulger D., Errington J.;

"Cloning and sequencing 7.5Kbp of DNA from Bacillus subtilis upstream

of the codv gene.";

RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

RL [2]

RP SEQUENCE OF 1-363 FROM N.A.

RC STRAIN=168 / 8G5;

RA de Jong S.;

"Cloning and sequencing of the TopI gene, the gene encoding B.

subtilis DNA topoisomerase I";

RT Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.

RL -1- SIMILARITY: BELONGS TO THE GIDA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

CC -----

DR EMBL: AJ000975; CAA04423.1; -

DR EMBL: Z99112; CAB13486.1; -

DR EMBL: L27797; AAA22764.1; -

DR Subtilist; BG11008; gid.

DR InterPro: IPR002218; GIDA.

DR InterPro: IPR00205; NAD binding.

DR PROSITE: PS01280; GIDA_1; FALSE_NEG.

DR PROSITE: PS01281; GIDA_2; 1.

KW Complete proteome.

SEQUENCE 435 AA; 48063 MW; 6C7AB028F484B683 CRC64;

OY 87 IIGAGIAGS 94

DB 9 IIGAGIAGS 16

Query Match 1.4%; Score 8; DB 1; Length 435;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4

ASTB_ECOLI

ID ASTB_ECOLI STANDARD; PRT; 447 AA.

AC P76216;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Succinylarginine dihydrolase (EC 3.-.-.-).

GN ASTB OR B1745.

OS Escherichia coli.


```

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
OC NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN (2)
RP FUNCTION.
RX MEDLINE=98361920; PubMed=9696779;
RA Schneider B.L., Knapakis A.K., Reitzer L.J.;
RT "Arginine catabolism and the arginine succinyltransferase pathway in
Escherichia coli.";
RT Bacteriol. 180:4278-4286(1998).
CC -1- FUNCTION: CATALYZES THE HYDROLYSIS OF N(2)-SUCCINYLRARGININE INFO
N(2)-SUCCINYLRORITHINE, AMMONIA AND CO(2).
CC -1- PATHWAY: SECOND STEP IN ARGININE CATABOLISM BY THE ARGININE
SUCCINYLRNTRANSFERASE PATHWAY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF000269; AAC74815.1; -
DR ECocore: EC13396; asbB.
KW Arginine metabolism; Hydrolase; Complete proteome.
SQ SEQUENCE 447 AA; 49298 MW; 710AE691E413AF64 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 447;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 EALDVLSSQ 103
Db 422 EALDVLSSQ 429

RESULT 5
IDS_ARATH STANDARD; PRT; 558 AA.
AC 038693; Q9LLYL; Q9CAV3;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)
DE (Carotene 7,8-desaturase).
GN ZDS1 OR ZDS OR AT3G04870 OR T9J14.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Scolnik P.A., Bartley G.E.;
RT "Nucleotide sequence of zeta-carotene desaturase from Arabidopsis.";
RL (in) Plant Gene Register PGR95-111.
RN (2)
RP SEQUENCE FROM N.A.
RA Giuliano G., Rosati C., Santangelo G.;
RT "Gene structure and regulation of the carotenoid biosynthetic pathway

```

```

RT in Arabidopsis.";
RN Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat B., Lemcke K., Rieger M., Ansoerge W., Unselid M.,
RA Farmanan M., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA Delenay M., Boulry M., Grivell L.A., Maché R., Puigdomenech P.,
RA de Simone V., Choisme N., Attiguenave F., Robert C., Brotier P.,
RA Wincker P., Catolico L., Weissenbach J., Saurin W., Queller F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
RA Reichelt J., Schafte M., Schoen O., Barques M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauró C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argilou A., Flores M., Iguori R., Vitale D.,
RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Wals A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Millscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nieman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasano T., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shindo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
thaliana.";
RL Nature 408:820-822(2000).
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
via the intermediary of neurosporene. It carries out two
consecutive desaturations (introduction of double bonds) at
positions C-7 and C-7'.
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neurosporene +
A + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Neurosporene + AH(2) + O(2) = lycopene + A + 2
H(2)O.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Chloroplast; Chromoplast.
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U38550; AAA91161.1; -
DR EMBL: AF121947; AAF85796.1; -
DR EMBL: AC009465; AAG51402.1; -
DR Mendel: 581; Arath; Zds1; 581.
DR InterPro: IPR000759; Adirndx_reductase.
DR InterPro: IPR000205; NAD_binding.
DR PRIN1: PR00419; ADXRPTASE.
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
KW Chloroplast; Transit peptide.
FT TRANSIT 1 27
FT CHAIN 28 558
FT CONFLICT 9 9 A -> P (IN REF. 1).
FT CONFLICT 93 93 G -> A (IN REF. 1).
FT CONFLICT 172 172 N -> S (IN REF. 1).
FT CONFLICT 370 370 G -> V (IN REF. 3).
FT CONFLICT 463 463 S -> P (IN REF. 1).
FT CONFLICT 470 470 C -> W (IN REF. 3).

```

SQ SEQUENCE 558 AA; 61508 MW; FDE5AE5DFB4A531D CRC64;

Query Match 1.4%; Score 8; DB 1; Length 558;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGLAG 93
| | | | | | | |
DB 61 IIGAGLAG 68

RESULT 6

ZDS_MAIZE
ID ZDS_MAIZE STANDARD; PRT; 570 AA.
AC Q9TPT4;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)
GN ZDS1 OR ZDS.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Rosaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo R., Wurtzel E.T.;
RT "A maize cDNA encoding zeta carotene desaturase."
RL (In) Plant Gene Register PGR99-118.
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
via the intermediary of neosporene. It carries out two
consecutive desaturations (introduction of double bonds) at
positions C-7 and C-7'.
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neosporene +
A + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Neosporene + AH(2) + O(2) = lycopene + A + 2
H(2)O.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AF047490; AAMD2462.1; -
DR Mendel; 35948; Zeama; zds1.35948.
DR InterPro: IPR000759; Adnrx_reductase.
DR InterPro: IPR003042; Rng_moxxygenase.
DR PRINTS; PR00419; AOXRDASE.
DR PRINTS; PR00420; RINGMOXGNASE.
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
KW Chloroplast; Transl. peptide.
FT TRANSIT 1 570 ZETA-CAROTENE DESATURASE.
FT CHAIN ? 570 ZETA-CAROTENE DESATURASE.
SQ SEQUENCE 570 AA; 63127 MW; C7ADB2FD62531B8 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 570;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGLAG 93
| | | | | | | |
DB 69 IIGAGLAG 76

RESULT 7

ZDS_NARPS
ID ZDS_NARPS STANDARD; PRT; 574 AA.
AC O49901;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)
GN ZDS1.
OS Narcissus pseudonarcissus (Daffodil).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
OC Narcissus.
OX NCBI_TaxID=39639;
RN [1]
RP SEQUENCE FROM N.A.
RA Al-Babili S., Oelischlegel J., Beyer P.;
RT "A cDNA encoding for beta carotene desaturase from Narcissus
pseudonarcissus L.";
RL (In) Plant Gene Register PGR98-103.
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
via the intermediary of neosporene. It carries out two
consecutive desaturations (introduction of double bonds) at
positions C-7 and C-7'.
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neosporene +
A + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Neosporene + AH(2) + O(2) = lycopene + A + 2
H(2)O.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AJ224683; CAAL2062.1; -
DR Mendel; 28308; Narps; zds1.28308.
DR InterPro: IPR002937; Amino-oxidase.
DR Pfam: PF01593; Amino-oxidase; 1.
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
KW Chloroplast; Transl. peptide.
FT TRANSIT 1 574 CHLOROPLAST (POTENTIAL).
FT CHAIN ? 574 ZETA-CAROTENE DESATURASE.
SQ SEQUENCE 574 AA; 63581 MW; 67D2C8DAFF942F77 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 574;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGLAG 93
| | | | | | | |
DB 69 IIGAGLAG 76

RESULT 8

ZDS_TARER
ID ZDS_TARER STANDARD; PRT; 587 AA.
AC Q9PTV46;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)
GN ZDS1.
OS Narcissus pseudonarcissus (Daffodil).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
OC Narcissus.
OX NCBI_TaxID=39639;
RN [1]
RP SEQUENCE FROM N.A.
RA Al-Babili S., Oelischlegel J., Beyer P.;
RT "A cDNA encoding for beta carotene desaturase from Narcissus
pseudonarcissus L.";
RL (In) Plant Gene Register PGR98-103.
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
via the intermediary of neosporene. It carries out two
consecutive desaturations (introduction of double bonds) at
positions C-7 and C-7'.
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neosporene +
A + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Neosporene + AH(2) + O(2) = lycopene + A + 2
H(2)O.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AJ224683; CAAL2062.1; -
DR Mendel; 28308; Narps; zds1.28308.
DR InterPro: IPR002937; Amino-oxidase.
DR Pfam: PF01593; Amino-oxidase; 1.
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
KW Chloroplast; Transl. peptide.
FT TRANSIT 1 574 CHLOROPLAST (POTENTIAL).
FT CHAIN ? 574 ZETA-CAROTENE DESATURASE.
SQ SEQUENCE 574 AA; 63581 MW; 67D2C8DAFF942F77 CRC64;

```

OS Tagetes erecta (African marigold).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
OC Asteridae: euasterids II: Asterales: Asteraceae: Asteroideae:
OC Helianthus: Tagetes.
OX NCBI_TaxID=13708;
RN [1]
RP SEQUENCE FROM N.A.
RA Moehs C.P., Tian L., Dellapenna D.;
RT "Analysis of carotenoid biosynthetic gene expression during marigold
RT petal development.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
CC via the intermediary of neurosporene. It carries out two
CC consecutive desaturations (introduction of double bonds) at
CC positions C-7 and C-7' (By similarity).
CC -!- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neurosporene +
CC A + 2 H(2)O.
CC -!- CATALYTIC ACTIVITY: Neurosporene + AH(2) + O(2) = lycopene + A + 2
CC H(2)O.
CC -!- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -!- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -!- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF251013; AAG10425.1; -
DR InterPro: IPR000759; Adrxndx_reductase.
DR InterPro: IPR003042; Rng_ammoxigenase.
DR PRINTS: PR00419; ADXRDTASE.
DR PRINTS: PR00420; RNMNMOXGNASE.
KM Carotenoid biosynthesis: Oxidoreductase; NAD; Flavoprotein; FAD;
KM Chloroplast; Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 587 ZETA-CAROTENE DESATURASE.
SQ SEQUENCE 587 AA; 65085 MW; 34D2DF150729B99C CRC64;

Query Match 1.4%; Score 8; DB 1; Length 587;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 ITGAGLAG 93
Db 87 ITGAGLAG 94

```

```

RA Albrecht M., Klein A., Hugueney P., Sandmann G., Kuntz M.;
RT "Molecular cloning and functional expression in E. coli of a novel
RT plant enzyme mediating zeta-carotene desaturation.";
RL FEBS Lett. 372:199-202(1995).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99421754; PubMed=10491195;
RA Albrecht M., Klein A., Hugueney P., Sandmann G., Kuntz M.;
RT "Catalytic properties of an expressed and purified higher plant type
RT zeta-carotene desaturase from Capsicum annuum.";
RL Eur. J. Biochem. 265:376-383(1999).
CC -!- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
CC via the intermediary of neurosporene. It carries out two
CC consecutive desaturations (introduction of double bonds) at
CC positions C-7 and C-7'.
CC -!- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neurosporene +
CC A + 2 H(2)O.
CC -!- CATALYTIC ACTIVITY: Neurosporene + AH(2) + O(2) = lycopene + A + 2
CC H(2)O.
CC -!- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -!- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -!- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X89897; CAA61985.1; -
DR InterPro: IPR000759; Adrxndx_reductase.
DR InterPro: IPR003042; Rng_ammoxigenase.
DR PRINTS: PR00419; ADXRDTASE.
DR PRINTS: PR00420; RNMNMOXGNASE.
KM Carotenoid biosynthesis: Oxidoreductase; NAD; Flavoprotein; FAD;
KM Chloroplast; Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 588 ZETA-CAROTENE DESATURASE.
SQ SEQUENCE 588 AA; 64684 MW; 55F568FAEE7EA91 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 588;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 ITGAGLAG 93
Db 84 ITGAGLAG 91

```

```

RESULT 9
ZDS_CAPAN
ID ZDS_CAPAN STANDARD; PRT; 588 AA.
AC Q9SMJ3;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, last sequence update)
DT 01-MAR-2002 (Rel. 41, last annotation update)
DE zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)
DE (Carotene 7,8-desaturase).
GN ZDS.
OS Capsicum annuum (Bell pepper).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
OC Asteridae: euasterids I; Solanales: Solanaceae: Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, LAMYO; TISSUE=fruit;
MEDLINE=96000207; PubMed=7556669;

```

```

RESULT 10
ZDS_LYCES
ID ZDS_LYCES STANDARD; PRT; 588 AA.
AC Q9SE20;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, last sequence update)
DT 01-MAR-2002 (Rel. 41, last annotation update)
DE zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)
DE (Carotene 7,8-desaturase).
GN ZDS.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
OC Asteridae: euasterids I; Solanales: Solanaceae: Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RA Bartley G.E., Ishida B.K.;
RT "Zeta-carotene desaturase from tomato.";

```

RL (In) Plant Gene Register PCR99-181.
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
CC via the intermediary of neurosporene. It carries out two
CC consecutive desaturations (introduction of double bonds) at
CC positions C-7 and C-7'.
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neurosporene +
CC A + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Neurosporene + AH(2) + O(2) = lycopene + A + 2
CC H(2)O.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF195507; AAF13698.1; -
CC InterPro: IPR000759; Adrnx_reductase.
CC InterPro: IPR003042; Rng_ammoxigenase.
CC PRINTS: PR00419; AOXRDASE.
CC PRINTS: PR00420; RINGMONOXINASE.
CC Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
CC Chloroplast; Transit peptide.
CC TRANSLIT ? 588 ZETA-CAROTENE DESATURASE.
CC CHAIN 1 ? CHLOROPLAST (POTENTIAL).
CC FT ? 588 ZETA-CAROTENE DESATURASE.
CC SEQUENCE 588 AA; 64733 MW; 786675800C1B846 CRC64;
CC -----
Query Match 1.4%; Score 8; DB 1; Length 588;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 86 IIGAGLAG 93
DB 84 IIGAGLAG 91

RESULT 11
FLDP_VIRBA STANDARD; PRT; 666 AA.
ID FLDP_VIRBA
AC Q56705;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polar flagellar hook-associated protein 2 (Haf2) (Filament cap
DE protein) (Flagellar cap protein).
DE FLDP OR FLAH.
DE N
DE Vibrio parahaemolyticus.
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
NCBI_TaxID=670;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB22;
RC MEDLINE=95189739; PubMed=7883718;
RA McCarter L.L.;
RT "Genetic and molecular characterization of the polar flagellum of
RT Vibrio parahaemolyticus."
RL J. Bacteriol. 177:1595-1609(1995).
CC -1- FUNCTION: REQUIRED FOR THE MORPHOGENESIS AND FOR THE ELONGATION OF
CC THE FLAGELLIN FILAMENT BY FACILITATING POLYMERIZATION OF THE
CC FLAGELLIN MONOMERS AT THE TIP OF GROWING FILAMENT. FORMS A CAPPING
CC STRUCTURE, WHICH PREVENTS FLAGELLIN SUBUNITS (TRANSPORTED THROUGH
CC THE CENTRAL CHANNEL OF THE FLAGELLUM) FROM LEAKING OUT WITHOUT
CC POLYMERIZATION AT THE DISTAL END. IMPORTANT FOR SWIMMING MOTILITY.
CC -1- SUBUNIT: HOMOPENTAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: FLAGELLAR.
CC -1- MISCELLANEOUS: V. PARAHAEVOLYTICUS POSSESSES TWO FLAGELLAR SYSTEMS;

CC A SINGLE POLAR FLAGELLUM PROPELS THE BACTERIUM IN LIQUID
CC (SWIMMING), WHILE MULTIPLE LATERAL (PERTRICHOUS) FLAGELLA MOVE
CC THE BACTERIUM OVER SURFACES (SWARMING). THE POLAR FLAGELLUM IS
CC SYNTHESIZED CONSTITUTIVELY BUT LATERAL FLAGELLA ARE PRODUCED ONLY
CC UNDER CONDITIONS IN WHICH THE POLAR FLAGELLUM IS NOT FUNCTIONAL.
CC -1- SIMILARITY: BELONGS TO THE FLID FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF069392; AAC27803.1; -
CC InterPro: IPR003481; FLID.
CC Pfam: PF02465; FLID.1.
CC Pfam: PF02465; FLID.1.
CC FLAGELLA; Coiled coil.
CC DOMAIN 338 424
CC SEQUENCE 666 AA; 72761 MW; EA2CA623ADA69F8A CRC64;
CC -----
Query Match 1.4%; Score 8; DB 1; Length 666;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 385 DPOERERL 392
DB 395 DPOERERL 402

RESULT 12
RIRL_BUCAI STANDARD; PRT; 761 AA.
ID RIRL_BUCAI
AC P57276;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonucleoside-diphosphate reductase alpha chain (EC 1.17.4.1)
DE (Ribonucleotide reductase).
DE NRDA OR BU179.
GN NRDA
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OS Bacteria; Proteobacteria; gamma subdivision; Buchnera.
NCBI_TaxID=118099;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RC MEDLINE=20445173; PubMed=10993077;
RC Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. Aps.;"
RL Nature 407:81-86(2000).
CC -1- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM
CC THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY
CC FOR DNA SYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
CC thiorodoxin + H(2)O = ribonucleoside diphosphate + reduced
CC thiorodoxin.
CC -1- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; AF001118; BABI2896.1; -.
DR InterPro; IPR000788; Ribonuclease_red.
DR Pfam; PF00317; ribonuc_red_1g; 1.
DR Pfam; PF02867; ribonuc_red_1gc; 1.
DR PRINTS; PR01183; RIBORDTASEM1.
DR PROSITE; PS00089; RIBORED_LARGE; 1.
KW Oxidoreductase; DNA replication; Complete proteome.
FT ACT_SITE 225 225 BY SIMILARITY.
FT ACT_SITE 439 439 BY SIMILARITY.
FT ACT_SITE 462 462 BY SIMILARITY.
FT SITE 754 754 INTERACTS WITH THIOREDOXIN
FT SITE 754 754 (BY SIMILARITY).
FT SITE 759 759 INTERACTS WITH THIOREDOXIN/
FT SITE 759 759 (BY SIMILARITY).
SQ SEQUENCE 761 AA; 87166 MW; 700A2A4AA426AC76 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 761;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 LAAKSAI 205
   |||||
Db 501 LAAKSAI 508

RESULT 13
GON2_DICLA STANDARD; PRT; 85 AA.
ID GON2_DICLA
AC Q91A08;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Gonadoliberin II precursor (Gonadotropin-releasing hormone II)
DE (GnRH-II) (LH-RH II) (Luliberin II).
GN GNRH2.
OS Dicertrarchus labrax (European sea bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;
OC Moronidae; Dicertrarchus.
OX NCBI_TaxID=13489;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Gonzalez-Martinez D., Madigou T., Zmora N., Anglade I., Zanny S.,
RA Zohar Y., Elizur A., Munoz-Cueto J.A., Kah O.;
RT "Differential expression of three different prepro-GnRH
RT (Gonadotropin-releasing hormone) messengers in the brain of the
RT European sea bass (Dicertrarchus labrax)".
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF224281; AAF62900.1; -.
DR InterPro; IPR002012; GNRH.
DR Pfam; PF00446; GNRH; 1.
DR PROSITE; PS00473; GNRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Signal.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 85 PROGONADOLIBERIN II.
FT CHAIN
SQ SEQUENCE 85 AA; 9673 MW; F832D53BC842C64 CRC64;

```

```

FT PEPTIDE 24 33 GONADOLIBERIN II.
FT PEPTIDE 37 85 GNRH-ASSOCIATED PEPTIDE II.
FT MOD_RES 24 24 PYROLIDONE CARBOXYLIC ACID
FT MOD_RES 33 33 (BY SIMILARITY).
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP) (BY
FT MOD_RES 33 33 SIMILARITY).
SQ SEQUENCE 85 AA; 9646 MW; F832C0698C842C64 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LFGLLC 14
   |||||
Db 9 LFGLLC 15

RESULT 14
GON2_MORSA STANDARD; PRT; 85 AA.
ID GON2_MORSA
AC Q73811;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Gonadoliberin II precursor (Gonadotropin-releasing hormone II)
DE (GnRH-II) (LH-RH II) (Luliberin II).
GN GNRH2.
OS Morone saxatilis (Striped bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;
OC Moronidae; Morone.
OX NCBI_TaxID=34816;
RN [1]
RP SEQUENCE FROM N.A.
RA Chow M.M., Kight K.E., Gotthilf Y., Alok D., Zohar Y.;
RT "Multiple GNRHs present in a teleost species are encoded by separate
RT genes: analysis of the sbgnrh and cgnrh-II genes from the striped
RT bass, Morone saxatilis."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF056313; AAD03816.1; -.
DR InterPro; IPR002012; GNRH.
DR Pfam; PF00446; GNRH; 1.
DR PROSITE; PS00473; GNRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Signal.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 85 PROGONADOLIBERIN II.
FT PEPTIDE 24 33 GONADOLIBERIN II.
FT PEPTIDE 37 85 GNRH-ASSOCIATED PEPTIDE II.
FT MOD_RES 24 24 PYROLIDONE CARBOXYLIC ACID
FT MOD_RES 33 33 (BY SIMILARITY).
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP) (BY
FT MOD_RES 33 33 SIMILARITY).
SQ SEQUENCE 85 AA; 9673 MW; F832D53BC842C64 CRC64;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LFGLLC 14
 |||||
 Db 9 LFGLLC 15

RESULT 15

VMEM_CVB ID VMEM_CVB STANDARD; PRT; 106 AA.
 AC P37989;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 11.4 kDa membrane protein (ORF 3).
 OS Chrysanthemum virus B (CVB).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
 OX NCBI_TaxID=12165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92013948; PubMed=1919520;
 RA LeMay K., Zavrlev S.;
 RT "Nucleotide sequence and gene organization of the 3'-terminal region
 of chrysanthemum virus B genomic RNA."
 RL J. Gen. Virol. 72:2333-2337(1991).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: TO ORF3 PROTEIN FROM POTEXVIRUSES AND TO THE 14 kDa
 CC PROTEIN FROM BSMV RNA 2BETA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S60150; AAB20078.2; -.
 DR PIR; JQ1248;
 DR InterPro; IPR001896; Plant_vir_prot.
 DR Pfam; PF01307; Plant_vir_prot; 1.
 DR ProDom; PD001561; Plant_vir_prot; 1.
 KW Transmembrane.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSHEM 72 92 POTENTIAL.
 SO SEQUENCE 106 AA; 11435 MW; B3366CFBB40C92D6 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 106;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 177 LSAATFL 183
 |||||
 Db 85 LSAATFL 91

Search completed: August 20, 2002, 04:30:39
 Job time: 278 sec

PCR primer used to
Human gene express
Human secreted pro
DNA encoding novel
Human cystine/Glu
Human EST-derived
Human lung tumour-
Human lung tumour-
PAD-CMV19, AAQ20
PAD-CMV19, express
PAD-CMV19, contig, a
PAD-CMV1, AAQ207
PAD-CMV1, express
Drosophila melanog
Vector pMVx-BG DNA
Human low adenosi
Human adenocarci re
Bacteriophage 44AH
Bacteriophage 44AH
cDNA encoding Hum
Colon tumour relat
Human immune/haema
Human immune/haema
Human immune/haema
Human neuroblastom
Human cDNA 5'-end
Human cDNA clone 1
Human neuroblastom
Human cDNA 5'-end
Human cDNA clone 1
Human neuroblastom
S. pneumoniae deriv
Human neuroblastom
Streptococcus pneu
DNA encoding a S.
Bacteriophage 44AH
DNA segment from p

DR WPI: 2001-244939/25.
 DR P-PSDB: AAT72972.

PT Novel Chlamydia pneumoniae outer membrane protein and polynucleotides
 PT encoding them, useful as components of vaccines for treating Chlamydia
 PT infections, and for detecting Chlamydia infections in the body fluids
 of mammals -
 XX
 XX
 PS
 PS

Claim 2; Fig 1; 82pp; English.

The present sequence is a DNA encoding Chlamydia pneumoniae OMP (outer membrane protein). OMP is useful in the production of vaccines, antibodies and pharmaceutical compositions which are useful for treating or preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum) infections. They are also useful as diagnostic reagents for detecting Chlamydia infection which involves assaying a body fluid of a mammal to be tested for the components. The OMP vaccine is useful in the preparation of a medicament for preventing and/or treating Chlamydia infection. The primers derived from OMP gene are also useful for detecting and/or identifying Chlamydia in a biological material. OMP antibodies are also useful as reagents for purifying OMP from a biological sample which involves carrying out antibody-based affinity chromatography with the biological sample. OMP gene is also useful in gene therapy.

Sequence 1907 BP; 542 A; 456 C; 380 G; 529 T; 0 other;

Query Match 100.0%; Score 1907; DB 22; Length 1907;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtggtgatttcaaaaaggccatgagtggttaattgaagtcctctacc 60
 Db 1 gtggtgatttcaaaaaggccatgagtggttaattgaagtcctctacc 60
 QY 61 aaacatgaaatctctagagagtggaatcgatctatgaggaacttcacatc 120
 Db 61 aaacatgaaatctctagagagtggaatcgatctatgaggaacttcacatc 120
 QY 121 tctcttgagcttatttctgtagtctccatctctctgttgtaattcccgagtc 180
 Db 121 tctcttgagcttatttctgtagtctccatctctctgttgtaattcccgagtc 180
 QY 181 tgaagtcataagatccttataatgaagtcgaatctacaagcagcgcttagcaacta 240
 Db 181 tgaagtcataagatccttataatgaagtcgaatctacaagcagcgcttagcaacta 240
 QY 241 tctggaagctctagatgctacggtgacatgacttctctgtttaagaaatcgagga 300
 Db 241 tctggaagctctagatgctacggtgacatgacttctctgtttaagaaatcgagga 300
 QY 301 agactatctcaagcaaaagcatcactcctcgaatccgcaactagaaagacacatcat 360
 Db 301 agactatctcaagcaaaagcatcactcctcgaatccgcaactagaaagacacatcat 360
 QY 361 tggagcaagcctggcgagatctcagaagccttgagcgtctccccaagctatggaac 420
 Db 361 tggagcaagcctggcgagatctcagaagccttgagcgtctccccaagctatggaac 420
 QY 421 tgcagacccctgcagacgactggtttatcgcagctcagagacatctgggaaac 480
 Db 421 tgcagacccctgcagacgactggtttatcgcagctcagagacatctgggaaac 480
 QY 481 ttctgacgactactgtttaagctttaagatcctctctctgcatccgcttagaagc 540
 Db 481 ttctgacgactactgtttaagctttaagatcctctctctgcatccgcttagaagc 540
 QY 541 cgcctatagcttgtaatttgaagaacactaaagtcattgatacctacattctcat 600
 Db 541 cgcctatagcttgtaatttgaagaacactaaagtcattgatacctacattctcat 600
 QY 601 tcaatagcttcccgagaatccaatgctactctgcgcaatatctcctaagcttagaac 660

Db 601 tcaatagcttcccgagaatccaatgctactctgcgcaatatctcctaagcttagaac 660
 QY 661 tgaagatctgagcttataattcggagatctctagctgcagaaagcgagatcgag 720
 Db 661 tgaagatctgagcttataattcggagatctctagctgcagaaagcgagatcgag 720
 QY 721 tgcacagctctgcagatcgagaatatacaaaaaagcctctctccgacacttagaa 780
 Db 721 tgcacagctctgcagatcgagaatatacaaaaaagcctctctccgacacttagaa 780
 QY 781 ttgctcaagagtggtctcctccaagatccaagaactatcttattgattaggaagct 840
 Db 781 ttgctcaagagtggtctcctccaagatccaagaactatcttattgattaggaagct 840
 QY 841 taagatgtgcagactactacaataaaaaagcaattgagaagcctgattgagatgt 900
 Db 841 taagatgtgcagactactacaataaaaaagcaattgagaagcctgattgagatgt 900
 QY 901 caattagacagatcgaagctttaaattgcttggggaaagagagacgctcttcggt 960
 Db 901 caattagacagatcgaagctttaaattgcttggggaaagagagacgctcttcggt 960
 QY 961 gataaaaaagcaagcacttgagagcggtcgagcctgagccctgtagcttagcattacc 1020
 Db 961 gataaaaaagcaagcacttgagagcggtcgagcctgagccctgtagcttagcattacc 1020
 QY 1021 ctctgagatagagatcctgagctcctgcgcatatctcctaaagcaagagcagc 1080
 Db 1021 ctctgagatagagatcctgagctcctgcgcatatctcctaaagcaagagcagc 1080
 QY 1081 caagtgaatgagctttagctctcttagagtgtaggtgtgacccctaaactactgga 1140
 Db 1081 caagtgaatgagctttagctctcttagagtgtaggtgtgacccctaaactactgga 1140
 QY 1141 atacattacggaagcctggtccacacatataatgagacttagccttagcttagcttc 1200
 Db 1141 atacattacggaagcctggtccacacatataatgagacttagccttagcttagcttc 1200
 QY 1201 taaggggctactttaaaaaattggaagcggtggaactatagcttagcttagcttc 1260
 Db 1201 taaggggctactttaaaaaattggaagcggtggaactatagcttagcttagcttc 1260
 QY 1261 ggaaggggaagagtggtgctccacaacccgagcctggaagagacatctcgttct 1320
 Db 1261 ggaaggggaagagtggtgctccacaacccgagcctggaagagacatctcgttct 1320
 QY 1321 ctccgctactcctaagaagcttaccctccctgatttaagaacttttgagagtcagaa 1380
 Db 1321 ctccgctactcctaagaagcttaccctccctgatttaagaacttttgagagtcagaa 1380
 QY 1381 aactcagcttgcacactgagatcttcttttaagtcacactcaacatcagaagcctt 1440
 Db 1381 aactcagcttgcacactgagatcttcttttaagtcacactcaacatcagaagcctt 1440
 QY 1441 agactcctttccaagctggaagcttctcctggaagacatcatccgcttagcaga 1500
 Db 1441 agactcctttccaagctggaagcttctcctggaagacatcatccgcttagcaga 1500
 QY 1501 tctgtcatttataactacacaaagatctcgaaaaaacgcttctccatcgatgata 1560
 Db 1501 tctgtcatttataactacacaaagatctcgaaaaaacgcttctccatcgatgata 1560
 QY 1561 aaaaaagcctaattcagaaacctgttatttggagcgggaaaaacaaagacccatcc 1620
 Db 1561 aaaaaagcctaattcagaaacctgttatttggagcgggaaaaacaaagacccatcc 1620
 QY 1621 cagcatgcccctacagcttatacagatcccgagaagcgttagcagatcattgttga 1680
 Db 1621 cagcatgcccctacagcttatacagatcccgagaagcgttagcagatcattgttga 1680
 QY 1681 tattctagagaactagcacactcgaagcttccgagaatatccgttattatgataact 1740

Db 1681 tatctagaagacactgaccctcgaagctctccgaagatattccgttatgatatacaact 1740
QY 1741 gatcagcgaagagagataaataattccagctctcagcagctactacataaattgt 1800
Db 1741 gatcagcgaagagagatgataaattccagctctcagcagctactacataaattgt 1800
QY 1801 ggaagtaaccccaactacgtctatgaagcttgctcttacttctagctctctgttc 1860
Db 1801 ggaagtaaccccaactacgtctatgaagcttgctcttacttctagctctctgttc 1860
QY 1861 ttatgggttctcgcgcgagcttgactatcaagagcctgtgaa 1907
Db 1861 ttatgggttctcgcgcgagcttgactatcaagagcctgtgaa 1907
RESULT 2
AAD03026/c
ID AAD03026 standard; DNA: 36 BP.
AC AAD03026;
XX
DT 13-JUN-2001 (first entry)
XX
DE Chlamydia pneumoniae OMP gene amplifying 3' PCR primer.
XX
KW Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;
KM antibody; medicament; Chlamydia infection; PCR primer; ss.
OS Chlamydia pneumoniae.
XX
PN WO200121804-A1.
XX
PD 29-MAR-2001.
XX
PF 15-SEP-2000; 2000WO-CA01088.
XX
PR 20-SEP-1999; 99US-0154652.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI: 2001-244939/25.
XX
PS Claim 32; Page 48; 82pp; English.
XX
CC The present sequence is a 3' PCR primer used for amplifying
CC Chlamydia pneumoniae OMP (outer membrane protein) gene. This primer
CC contains a Bam HI restriction site and the sequence encoding the
CC C-terminal sequence of OMP. OMP is useful in the production of vaccines,
CC antibodies and pharmaceutical compositions which are useful for treating
CC or preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or
CC C. pecorum) infections. They are also useful as diagnostic reagents for
CC detecting Chlamydia infection which involves assaying a body fluid
CC of a mammal to be tested for the components. The OMP vaccine is
CC useful in the preparation of a medicament for preventing and/or
CC treating Chlamydia infection. The primers derived from OMP gene are
CC also useful for detecting and/or identifying Chlamydia in a biological
CC material. OMP antibodies are also useful as reagents for purifying OMP
CC from a biological sample which involves carrying out antibody-based
CC affinity chromatography with the biological sample. OMP gene is
CC also useful in gene therapy.
XX
SQ Sequence 36 BP; 8 A; 12 C; 7 G; 9 T; 0 other;

Query Match 1.3%; Score 24; DB 22; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.071;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1781 ggcctactcacaataatgtgag 1804
Db 36 GGCTTACTCATATAAATGTGTGAG 13
RESULT 3
AAD03025
ID AAD03025 standard; DNA: 44 BP.
AC AAD03025;
XX
DT 13-JUN-2001 (first entry)
XX
DE Chlamydia pneumoniae OMP gene amplifying 5' PCR primer.
XX
KW Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;
KM antibody; medicament; Chlamydia infection; PCR primer; ss.
OS Chlamydia pneumoniae.
XX
PN WO200121804-A1.
XX
PD 29-MAR-2001.
XX
PF 15-SEP-2000; 2000WO-CA01088.
XX
PR 20-SEP-1999; 99US-0154652.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI: 2001-244939/25.
XX
PS Claim 32; Page 48; 82pp; English.
XX
CC The present sequence is a 5' PCR primer used for amplifying
CC Chlamydia pneumoniae OMP (outer membrane protein) gene. This primer
CC contains a Not I restriction site, a ribosome binding site, an
CC initiation codon and a sequence at the 5' end of the OMP coding
CC sequence. OMP is useful in the production of vaccines, antibodies
CC and pharmaceutical compositions which are useful for treating or
CC preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or
CC C. pecorum) infections. They are also useful as diagnostic reagents for
CC detecting Chlamydia infection which involves assaying a body fluid
CC of a mammal to be tested for the components. The OMP vaccine is
CC useful in the preparation of a medicament for preventing and/or
CC treating Chlamydia infection. The primers derived from OMP gene are
CC also useful for detecting and/or identifying Chlamydia in a biological
CC material. OMP antibodies are also useful as reagents for purifying OMP
CC from a biological sample which involves carrying out antibody-based
CC affinity chromatography with the biological sample. OMP gene is
CC also useful in gene therapy.
XX
SQ Sequence 44 BP; 12 A; 14 C; 8 G; 10 T; 0 other;

Query Match 1.3%; Score 24; DB 22; Length 44;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 101 atggagactatccatctactctc 124
Db 21 atggagactatccatctactctc 44

AA75007	4	RESULT	
ID	AAA75007	standard; DNA: 4343 BP.	
XX	AAA75007;		
XX			
DT	02-JAN-2001	(first entry)	
DE	Nucleotide sequence of an Influenza virus polynucleotide.		
XX			
KW	Recombinant virus; vaccine; infection; hepatitis C virus vaccine;		
KW	human immunodeficiency virus vaccine; HIV vaccine; tumour vaccine; ss.		
OS	Influenza virus.		
XX			
PN	EP1035209-A1.		
PD	13-SEP-2000.		
XX			
PF	06-MAR-1999;	99EP-0104519.	
XX			
PR	06-MAR-1999;	99EP-0104519.	
XX			
PA	(ARTE-) ARTEMIS PHARM GMBH.		
XX			
PI	Hobom G, Flick R, Menke A, Azzey M;		
DR	WPI; 2000-559876/52.		
XX			
PT	Recombinant influenza virus useful for gene therapy and as a vaccine		
PT	against influenza and other infections, comprises viral RNA molecule		
PT	exchanged for a viral RNA encoding foreign gene		
PS	Disclosure; Page 22-24; 49pp; English.		
XX			
CC	The specification describes a recombinant influenza virus, which is		
CC	genetically stable in the absence of helper virus. The influenza virus		
CC	has at least one of the regular viral RNA segments exchanged for a		
CC	viral RNA encoding a foreign gene. The virus also has at least one of		
CC	the regular viral RNA segments in an antisense RNA molecule containing		
CC	one of the standard viral genes in sense orientation and a foreign,		
CC	recombinant gene in anti-sense orientation, or vice versa. The		
CC	recombinant viruses are useful for preparing vaccines in the form of		
CC	inactivated preparations or live recombinant viruses which are protective		
CC	against influenza and against other infections. The viruses are also		
CC	useful in somatic gene therapy, for transfer and expression of foreign		
CC	genes into cells infected by such viruses. They are also useful for		
CC	design and production of hepatitis C virus, human immunodeficiency		
CC	virus (HIV) and tumour vaccines. The present sequence represents an		
CC	influenza virus sequence which is used to produce viruses of the		
CC	invention.		
XX			
SQ	Sequence 4343 BP; 1085 A; 1066 C; 1100 G; 1092 T; 0 other:		
	Query Match	1.1%; Score 21; DB 21; Length 4343;	
	Best Local Similarity	100.0%; Pred. No. 2.7;	
	Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1546 tctccatgattatgcacaaa 1566		
TD	1876 tctccatgattatgcacaaa 1896		
	RESULT 5		
ID	AA94540		
XX	AA94540 standard; DNA: 20 BP.		
AC	AA94540;		
XX			
DT	13-SEP-1999	(first entry)	
DE	PCR primer used to amplify an ORF of Chlamydia pneumoniae.		

XX	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KM	vaccine; neutralising epitope; PCR primer; ss.
XX	Synthetic.
OS	Chlamydia pneumoniae.
PN	WO927105-A2.
XX	
PD	03-JUN-1999.
XX	
PF	20-NOV-1998; 98WO-IB01890.
XX	
PR	04-NOV-1998; 98US-0107078.
XX	21-NOV-1997; 97FR-0014673.
PA	(GEST) GENSET.
XX	
PI	Griifals R;
DR	WPI, 1999-357842/30.
XX	
PT	Genome sequence of Chlamydia pneumoniae
XX	
PS	Page 1678; Disclosure; 1912pp; English:
CC	AAx91991-x97517 represent PCR primers used to amplify open reading
CC	frames and other nucleic acid sequences from the genome of
CC	Chlamydia pneumoniae (see AAx91990). C. pneumoniae causes respiratory
CC	disease such as pneumonia and bronchitis and is thought to be a
CC	contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC	otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC	by the open reading frames of the C. pneumoniae genome (see AAT4584-
CC	AAx35879) can be used in immunogenic compositions as vaccines. Vectors
CC	containing C. pneumoniae nucleotide sequences can also be used as
CC	immunogenic compositions, especially where the vector directs the
CC	expression of a neutralising epitope of C. pneumoniae.
XX	
SQ	Sequence 20 BP; 6 A; 4 C; 6 G; 4 T; 0 other:
Query Match	1.0%; Score 20; DB 20; Length 20;
Best Local Similarity	100.0%; Pred. No. 8.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	844 ggatgctcagagctactaca 863
DB	1 ggatgctcagagctactaca 20
RESULT 6	
AAx94538	ID AAx94538 standard; DNA: 20 BP.
AC	AAx94538;
DT	13-SEP-1999 (first entry)
DE	PCR primer used to amplify an ORF of Chlamydia pneumoniae.
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KM	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KX	vaccine; neutralising epitope; PCR primer; ss.
OS	Synthetic.
OS	Chlamydia pneumoniae.
PN	WO9927105-A2.
XX	
PD	03-JUN-1999.
PF	20-NOV-1998; 98WO-IB01890.

XX 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97PR-0014673.
 XX (GEST) GENSET.
 PA Griffiths R;
 PI Griffiths R;
 DR WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PS Page 1677; Disclosure; 1912pp; English.
 XX
 CC AAX91991-X97517 represent PCR primers used to amplify open reading
 CC frames and other nucleic acid sequences from the genome of
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
 CC disease such as pneumonia and bronchitis and is thought to be a
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
 CC containing C. pneumoniae nucleotides sequences can also be used as
 CC immunogenic compositions, especially where the vector directs the
 CC expression of a neutralising epitope of C. pneumoniae.
 CC
 SQ Sequence 20 BP; 3 A; 6 C; 3 G; 8 T; 0 other;

Query Match 1.0%; Score 20; DB 20; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1541 cgtctctcatgattatgc 1560
 DB 1 cgtctctcatgattatgc 20

RESULT 7
 ID AAX92063/C
 XX AAX92063 standard; DNA; 20 BP.
 AC AAX92063;
 XX
 DT 13-SEP-1999 (first entry)
 DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; PCR primer; ss.
 XX
 OS Synthetic.
 OS Chlamydia pneumoniae.
 XX
 PN WO9927105-A2.
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97PR-0014673.
 XX
 PA (GEST) GENSET.
 PI Griffiths R;
 DR WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PS Page 1482; Disclosure; 1912pp; English.

XX AAX91991-X97517 represent PCR primers used to amplify open reading
 CC frames and other nucleic acid sequences from the genome of
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
 CC disease such as pneumonia and bronchitis and is thought to be a
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
 CC containing C. pneumoniae nucleotides sequences can also be used as
 CC immunogenic compositions, especially where the vector directs the
 CC expression of a neutralising epitope of C. pneumoniae.
 CC
 SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;

Query Match 1.0%; Score 20; DB 20; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1288 ccgaggtcttgaagacaga 1307
 DB 20 CCGAGGTCTTGAAGACAGA 1

RESULT 8
 ID AAX92065/C
 XX AAX92065 standard; DNA; 20 BP.
 AC AAX92065;
 XX
 DT 13-SEP-1999 (first entry)
 DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; PCR primer; ss.
 XX
 OS Synthetic.
 OS Chlamydia pneumoniae.
 XX
 PN WO9927105-A2.
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97PR-0014673.
 XX
 PA (GEST) GENSET.
 PI Griffiths R;
 DR WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PS Page 1482; Disclosure; 1912pp; English.
 XX
 CC AAX91991-X97517 represent PCR primers used to amplify open reading
 CC frames and other nucleic acid sequences from the genome of
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
 CC disease such as pneumonia and bronchitis and is thought to be a
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
 CC containing C. pneumoniae nucleotides sequences can also be used as
 CC immunogenic compositions, especially where the vector directs the
 CC expression of a neutralising epitope of C. pneumoniae.
 CC

Sequence 20 BP; 4 A; 2 C; 10 G; 4 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 20; DB 20; Length 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 319 catcacctccatccgacgc 338
|||||

DB 20 CATCCACTCTCCAGATCCGC 1

RESULT 9

AAx92068/c
ID AAX92068 standard; DNA; 20 BP.

AC AAX92068;

DT 13-SEP-1999 (first entry)

DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; PCR primer; ss.

OS Synthetic.

OS Chlamydia pneumoniae.

PN MO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

PI Griffiths R;

DR WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

PS Page 1482; Disclosure; 1912pp; English.

XX AAX91991-X97517 represent PCR primers used to amplify open reading
frames and other nucleic acid sequences from the genome of

CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
disease such as pneumonia and bronchitis and is thought to be a

CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
otitis media, erythema nodosum or pharyngitis. The polypeptides encoded

CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
AAY35879) can be used in immunogenic compositions as vaccines. Vectors

CC containing C. pneumoniae nucleotide sequences can also be used as
immunogenic compositions, especially where the vector directs the

CC expression of a neutralising epitope of C. pneumoniae.

XX Sequence 20 BP; 1 A; 2 C; 8 G; 9 T; 0 other;

Query Match 1.0%; Score 20; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1601 gaaacccaagaccatcc 1620
|||||

DB 20 GAAACCAAGACCCCATCC 1

RESULT 10

AAx92070/c
ID AAX92070 standard; DNA; 20 BP.

AC AAX92070;

DT 13-SEP-1999 (first entry)

DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; PCR primer; ss.

OS Synthetic.

OS Chlamydia pneumoniae.

PN MO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

PI Griffiths R;

DR WPI; 1999-357842/30.

PT Genome sequence of Chlamydia pneumoniae

PS Page 1483; Disclosure; 1912pp; English.

XX AAX91991-X97517 represent PCR primers used to amplify open reading
frames and other nucleic acid sequences from the genome of

CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
disease such as pneumonia and bronchitis and is thought to be a

CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
otitis media, erythema nodosum or pharyngitis. The polypeptides encoded

CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
AAY35879) can be used in immunogenic compositions as vaccines. Vectors

CC containing C. pneumoniae nucleotide sequences can also be used as
immunogenic compositions, especially where the vector directs the

CC expression of a neutralising epitope of C. pneumoniae.

XX Sequence 20 BP; 4 A; 4 C; 8 G; 4 T; 0 other;

Query Match 1.0%; Score 20; DB 20; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1272 ggttgcttcacaccca 1291
|||||

DB 20 GGTGCTCTCCACCAACCGA 1

RESULT 11

AAZ16609
ID AAZ16609 standard; cDNA; 800 BP.

AC AAZ16609;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:4079.

KW Human; gene; gene expression product; diagnosis; therapy; probe;
detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.
 XX
 XX MO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Ctkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia A, Garcia V, Glese K, Inlis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Suduth-Klinger J, Williams LT;
 DR WPI; 1999-494092/41.
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX
 PS Claim 1: Page 1934; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA217779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 CC
 XX Sequence 800 BP; 199 A; 166 C; 202 G; 211 T; 22 other;
 SQ

Query Match 1.0%; Score 19; DB 20; Length 800;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 351 gcacatcattggagcagg 369
 ||||||||||||||||
 DB 500 gcacatcattggagcagg 518

RESULT 12
 AAD07852 standard; cDNA; 1210 BP.
 ID AAD07852 standard; cDNA; 1210 BP.
 XX
 AC AAD07852;
 XX
 DT 10-AUG-2001 (first entry)
 XX
 DE Human secreted protein-encoding gene 6 cDNA clone HFTF82, SEQ ID NO:54.

XX
 KW Human: secreted protein; proliferative disorder; cancer; tumour; asthma;
 KW foetal abnormality; developmental abnormality; hematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;
 KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KW cell culture; chemotaxis; vulnery; binding partner identification;
 KW gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT complement (770..1129)
 FT CDS
 FT /tag- a
 FT /product= "Human secreted protein precursor"
 FT /note= "CDS does not include start codon"
 FT /partial
 FT sig-peptide complement (1127..1129)
 FT /tag- b
 FT mat-peptide complement (773..1126)
 FT /tag- c
 FT /product= "Mature human secreted protein"
 PN WO200132675-A1.
 XX
 PD 10-MAY-2001.
 XX
 PF 25-OCT-2000; 2000WO-US29363.
 XX
 PR 29-OCT-1999; 99US-0162239.
 PR 30-JUN-2000; 2000US-0215139.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Komatsoulis GA, Wei P, Baker KP, Young PE;
 DR WPI; 2001-338772/34.
 DR P-PSDB; AAE03470.
 XX
 PT Thirty two human secreted proteins, useful for treating cancers,
 PT hyperproliferative disorders, inflammatory disorders, neurological
 PT disorders, autoimmune diseases and cardiovascular disorders -
 XX
 PS Claim 1: Page 438; 576pp; English.
 XX
 CC AAD07809-AAD07907 represent cDNAs corresponding to 32 human secreted
 CC protein genes, and AAE03427-AAE03523 represent the proteins they encode.
 CC AAE03524-AAE03537 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 32 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC hematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angioenic disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated

CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein-encoding cDNA of
CC the invention.
XX
XX
SQ Sequence 1210 BP; 307 A; 337 C; 285 G; 278 T; 3 other;

Query Match 1.0%; Score 19; DB 22; Length 1210;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 750 aacaaacgccttcttc 768
Db 282 aacaaacgccttcttc 300

RESULT 13
ID AAS64487 standard; cDNA: 1665 BP.
XX
XX AAS64487;
AC
XX
DT 13-FEB-2002 (first entry)
DE
XX DNA encoding novel human diagnostic protein #291.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR
XX P-PSDB: ABG00300.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX
XX Claim 1; SEQ ID No 291; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes.
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp:wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1665 BP; 386 A; 497 C; 463 G; 319 T; 0 other;

Query Match 1.0%; Score 19; DB 23; Length 1665;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 931 ttggggaaagaagagac 949
Db 1291 ttggggaaagaagagac 1309

RESULT 14
ID ABA09201 standard; cDNA: 1958 BP.
XX
XX ABA09201;
AC
XX
DT 11-JAN-2002 (first entry)
DE
XX Human cystine/Glu transporter homologue cDNA, SEQ ID NO:977.
XX
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; anti-inflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antilucer; ss.
XX
XX Homo sapiens.
XX
XX WO200157188-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US03800.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-457740/49.
DR
XX P-PSDB: ABB11957.
XX
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
XX
XX Claim 1; Page 835; 1963pp; English.
XX
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides,
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence

CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haemopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention.

Sequence 1958 BP; 499 A; 418 C; 424 G; 617 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 1958;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 351 gcaccatcattggagcagg 369
Db 226 gcaccatcattggagcagg 244
|||||

RESULT 15

AAH98573 standard; cDNA; 1958 BP.

AAH98573;

12-OCT-2001 (first entry)

Human EST-derived coding sequence SEQ ID NO: 430.

Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

diagnostics; forensic test; gene mapping; genetic disorder;

biodiversity; gene therapy; nutrition; ss.

Homo sapiens.

WO200154477-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US02687.

25-JAN-2000; 2000US-0491404.

17-JUL-2000; 2000US-0617746.

03-AUG-2000; 2000US-0631451.

15-SEP-2000; 2000US-0663870.

(HYSE-) HYSEQ INC.

PI

XX WPI; 2001-476164/51.
DR P-PSDB; AAM23914.
XX
PT isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 484; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.

Sequence 1958 BP; 499 A; 418 C; 424 G; 617 T; 0 other;
Query Match 1.0%; Score 19; DB 22; Length 1958;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 351 gcaccatcattggagcagg 369
Db 226 gcaccatcattggagcagg 244
|||||

Search completed: August 20, 2002, 03:16:11
Job time: 4616 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 03:20:55 ; Search time 42.94 Seconds
(without alignments)
1271.047 Million cell updates/sec

Title: US-09-662-812-2
Perfect score: 2833
Sequence: 1 MGLFHLTFGLLCSLPSL.....TGDAKNFPVLAKLTIKIVE 568

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR-71:*
1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2833	100.0	568	2 C72129	probable outer mem
2	2833	100.0	568	2 G86493	probable leader pe
3	1456	51.4	566	2 A81682	conserved hypochet
4	1421	50.2	566	2 G71525	probable outer mem
5	203	7.2	1547	2 A12043	hypothetical prote
6	164	5.8	1119	2 AB2239	hypothetical prote
7	153	5.4	1348	2 AG2558	hypothetical prote
8	148.5	5.2	2297	2 AB2494	hypothetical prote
9	145	5.1	1010	2 AH2553	hypothetical prote
10	145	5.1	1381	2 AF2010	regulatory protein
11	134.5	4.7	730	2 S28294	hypothetical prote
12	134.5	4.7	826	2 C88550	hypothetical prote
13	134.5	4.7	1238	1 S17944	protein ZC84.3 lim
14	134.5	4.7	1238	1 S17946	virulence sensor p
15	134	4.7	860	2 T37768	virulence sensor p
16	133	4.7	321	2 F83358	probable vacuolar
17	133	4.7	1069	2 AF1930	hypothetical prote
18	133	4.7	1138	2 G71534	hypothetical prote
19	131.5	4.6	2469	2 H36812	probable transmemb
20	131	4.6	398	2 T44927	hypothetical prote
21	131	4.6	398	2 AD1882	hypothetical prote
22	129.5	4.6	2337	2 T40577	hypothetical prote
23	128	4.5	1189	2 A54817	ATPase Scitl, chrom
24	127	4.5	1238	1 A40185	probable protein
25	126.5	4.5	1292	2 D84727	virulence protein
26	126	4.4	631	2 JCA298	hyaluronan recepto
27	126	4.4	789	2 T38423	hypothetical prote
28	126	4.4	851	2 AG2469	hypothetical prote
29	125	4.4	1805	1 A64224	hypothetical prote

30	124.5	4.4	526	2 B81679	conserved hypochet
31	124	4.4	1275	2 T49362	hypothetical prote
32	124	4.4	1708	2 AE1866	WD-40 repeat prote
33	123.5	4.4	894	2 T13029	beta-adaptin homol
34	122	4.3	719	2 T05384	beta-adaptin homol
35	121.5	4.3	556	2 B46024	neurofilament-L su
36	121.5	4.3	866	2 A11880	aminopeptidase lim
37	121.5	4.3	2555	2 C69681	peptide synthetase
38	121.5	4.3	566	2 G75457	tetratricopeptide
39	121	4.3	1603	1 VUKWS	vitellinogen vit-5
40	121	4.3	1603	2 F89497	protein vit-5 (imp
41	120.5	4.3	430	2 F75053	hypothetical prote
42	120.5	4.3	937	2 A35553	beta-adaptin - hum
43	120.5	4.3	937	2 C35553	beta-adaptin - rat
44	120	4.2	922	2 G83109	probable two-compo
45	119.5	4.2	1466	2 G84516	probable retroelem

ALIGNMENTS

RESULT	1	
C72129	1	
Probable outer membrane leader peptide (omp) CP0021 - Chlamydia pneumoniae (str		
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae		
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000		
C:Accession: C72129, B81542		
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,		
Nature Genet. 21, 385-389, 1999		
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.		
A:Reference number: A72000; MUID:99206606		
A:Accession: C72129		
A:Molecule type: DNA		
A:Residues: 1-568 <ARN>		
A:Cross-references: GB:AE001587; GB:AE001363; NID:94376271; PIDN:AD18174.1; PID:943		
A:Experimental source: strain CML029		
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hick		
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salz		
Nucleic Acids Res. 28, 1397-1406, 2000		
A:Title: Genome sequences of Chlamydia trachomatis MoRa and Chlamydia pneumoniae AR3		
A:Reference number: A81500; MUID:20150255		
A:Accession: B81542		
A:Molecule type: DNA		
A:Residues: 1-568 <REA>		
A:Cross-references: GB:AE002234; GB:AE002161; NID:97189667; PIDN:AF38558.1; PID:971		
A:Experimental source: strain AR39, HL cells		
C:Genetics:		
A:Gene: CP0021; CP0755		
Query Match		
Best Local Similarity 100.0%; Score 2833; DB 2; Length 568;		
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	MGLFHLTFGLLCSLPSLVAKFPESVGHILVISTOSTOQALATYLEALDAYGDHDF 60
DB	1	MGLFHLTFGLLCSLPSLVAKFPESVGHILVISTOSTOQALATYLEALDAYGDHDF 60
QY	61	VLRKIGEDYLRKQSHSSDPQTRKSTIGAGLAGSSEALDVLSQMETADPIQQLVLVSAY 120
DB	61	VLRKIGEDYLRKQSHSSDPQTRKSTIGAGLAGSSEALDVLSQMETADPIQQLVLVSAY 120
QY	121	SGHLKTSDDLFRALASPYVIRLEAAYRLANKTKVVIDHLHSFHKRLPEEIOCLSA 180
DB	121	SGHLKTSDDLFRALASPYVIRLEAAYRLANKTKVVIDHLHSFHKRLPEEIOCLSA 180
QY	181	IFLRLEPESDAYIRDLAANKSAIRSATALQIGGYOQKRLPTLRNLDTSPDOQBAI 240
DB	181	IFLRLEPESDAYIRDLAANKSAIRSATALQIGGYOQKRLPTLRNLDTSPDOQBAI 240
QY	241	LYALGKLKDGSGYNIKKQLOKPDVDTVLAQAOLIALGKEDALPVIKKQALERRPAL 300
DB	241	LYALGKLKDGSGYNIKKQLOKPDVDTVLAQAOLIALGKEDALPVIKKQALERRPAL 300

QY	301	YALRRPSEIGIPLIPFLFKTKNSEAKLNVALALLEGCOTPKLLEYITTELVOPHINE	360
Db	301	YALRRLPSEIGIPLIPFLFKTKNSEAKLNVALALLEGCOTPKLLEYITTELVOPHINE	360
QY	361	TLALSFSGKRTLQNMKNRYNIIVPODPOHERERLLSTTGLSEEOILTFELFRLPREAVALPCY	420
Db	361	TLALSFSGKRTLQNMKNRYNIIVPODPOHERERLLSTTGLSEEOILTFELFRLPREAVALPCY	420
QY	421	KLASOKQOLMTATISFSLTSHOALDLFLQAAALPGEPIIRAYADAIYVLTWDPPEK	480
Db	421	KLASOKQOLMTATISFSLTSHOALDLFLQAAALPGEPIIRAYADAIYVLTWDPPEK	480
QY	481	RSIHHYAKKLIQIETLLFVDTENQRHPMPMYLRYOVTPESRTKMLMDIETLATSXSSED	540
Db	481	RSIHHYAKKLIQIETLLFVDTENQRHPMPMYLRYOVTPESRTKMLMDIETLATSXSSED	540
QY	541	IRLLIQLMTEGDAKNFPVLACILIKTIVE	568
Db	541	IRLLIQLMTEGDAKNFPVLACILIKTIVE	568

```

RESULT      2
GB6493
probable leader peptide Omp [imported] - Chlamydomophila pneumoniae (strain J138)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: GB6493
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; I
Nucleic Acids Res. 28, 2311-2314, 2000
A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: AB6491; MUID:20330349
A:Accession: GB6493
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <STO>
A:Cross-references: GB:BA000008; NID:98978396; PIDN:BAA98233.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
:Gene: CPJ0021

```

Query Match	Similarity	100.0%	Score 2833;	DB 2;	Length 568;
Best Local	Similarity	100.0%	Pred. No. 4,5e-164;		
Matches	568;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1	MGFLHFLTGFLGCLSPISLVAKPEPSYGHKLLYSTOSTQOALATYLEALDAYGDHDF	60		
Db	1	MGFLHFLTGFLGCLSPISLVAKPEPSYGHKLLYSTOSTQOALATYLEALDAYGDHDF	60		
QY	61	VLKRGEDYLNKSHSSPQTRKSTIIAGAGLSSSEALDYSQAMETADPLQOLLVLSAV	120		
Db	61	VLKRGEDYLNKSHSSPQTRKSTIIAGAGLSSSEALDYSQAMETADPLQOLLVLSAV	120		
QY	121	SGHLKTSDDLFLKALASPVIYRIEAAVYRIANLKNTRYVIDHLHSFIKRLPEIOCLSA	180		
Db	121	SGHLKTSDDLFLKALASPVIYRIEAAVYRIANLKNTRYVIDHLHSFIKRLPEIOCLSA	180		
QY	181	IFLRLTEESDVAIRDLIAAKKSATIRSATATQIGEYQOKREPLTLRNLLTSASPODOEAI	240		
Db	181	IFLRLTEESDVAIRDLIAAKKSATIRSATATQIGEYQOKREPLTLRNLLTSASPODOEAI	240		
QY	241	LYALGKLKDGOSYNIKKQLOKPDVDVYLAQAALIAIGKEDDALPVYIKKQALERRPAL	300		
Db	241	LYALGKLKDGOSYNIKKQLOKPDVDVYLAQAALIAIGKEDDALPVYIKKQALERRPAL	300		
QY	301	YALRHLPSEIGIPIALPIFLKTKNSEAKLVNALLEIGCDTPKLLIEXITERLVQPHNE	360		
Db	301	YALRHLPSEIGIPIALPIFLKTKNSEAKLVNALLEIGCDTPKLLIEXITERLVQPHNE	360		
QY	361	TLALSFSGRGLQNMWKRNIIVPODPQERKLLSTTRKLEOILTFELRLPREKAVLPCTY	420		
Db	361	TLALSFSGRGLQNMWKRNIIVPODPQERKLLSTTRKLEOILTFELRLPREKAVLPCTY	420		

Qy	421	KLLAQKQQLATTAISFSLSHSOEDLDFQAAKLPGEPIRAYADLAIYNLTQDEPK	480
Db	421	KLLAQKQQLATTAISFSLSHSOEDLDFQAAKLPGEPIRAYADLAIYNLTQDEPK	480
Qy	481	RSLSHYAKKLQETLLFVDTENQRHPSMPYLRYQVTPESSTKMLDILETLATSKSSED	540
Db	481	RSLSHYAKKLQETLLFVDTENQRHPSMPYLRYQVTPESSTKMLDILETLATSKSSED	540
Qy	541	IRLLQLMTSGDAKNFVLAGLLIKIYE	568
Db	541	IRLLQLMTSGDAKNFVLAGLLIKIYE	568

RESULT 3
A:1682
conserved hypothetical protein TC0629 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: A81682
R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey
C: Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR399
A:Reference number: A81500; MUID:20150255
A:Accession: A81682
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-566 <TEP>
A:Cross-references: GB:AE002331; GB:AE002160; NID:g7190662; PIDN:AAF39458.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0629

	Query Match	51.4%	Score 1456;	DB 2;	Length 566;
	Best Local Similarity	53.5%;	Pfed. No. 1;le-80;		
	Matches 304;	Conservative	92;	Mismatches 170;	Indels 2;
				Gaps	1;
OY	1 MGLFHLTFGLLLCSLPISLVAKFPESVGHKLLYSTOSTOQALATYTFEALDAGCDHFF	60			
Db	1 MGLSLLILFGLL -SLPISASDCDFPSVSOKLFLFLCQKSIPGALSYLEASTYYQQHNS	58			
OY	61 VLRKTGEYLKOSIHSSDPOTRKSTTIGAGLAGSSEBALDVLSQAMETADPLQOLLVLVAS	120			
Db	59 ILRLAKSYLOQSLESDAAYVKRKAIIAGLGSSESETDLDSESIETODLYEQOLLITLNA	118			
OY	121 SGHIGTSDDLLFEFKLASPYPIRYIREAAVIRLANENKVYIDHLHSFIHKLRPEIOLCSA	180			
Db	119 GNQJCKTSDRLLFKLTZPHPTIRILEAAVIRLAMCNKSKVSDLYEFTHQLPEIYNLAAT	178			
OY	181 IFLRETESDAYIRDLLIAANKSAIRSATAIOIGEYOQRRLPTLRNLLTSASPDOBAI	240			
Db	179 IFLEETEADAYAIRLLSSPNSLTRNYMAYLIGEGORRFPLTRLRSLLTSAAPLDQGS	238			
OY	241 LVALGKLTDGSGSYNIKKOLOKPDVDYTLAAQAIALIGKEDDALPVTKKQALERPAL	300			
Db	239 LYAIKGLTEDASSYPRIKALKSSKSNEVAALAAQTLLFGKDEALPILTTCOOQLPRAI	298			
OY	301 YALRHLPSEIGIPILAIFLFTKKNSEAKLNVALALLGECDTPKLEYETIRLVOPHNE	360			
Db	299 YTSRSLSEKKEBELLRPFCFAIKKEIKLNMAALNVHGSVNHVLVSLTEFLENKILHR	358			
OY	361 TLVLSFSKGRTPLQNMKRVIIVPODERERLLSTTRGLEQILLFERLPKEAVLPCITY	420			
Db	359 IFLPTHSIGKATOWKECTALPLRLSPBEKARALAMYRAAEPTILSSLTKLNPNNAALPYLE	418			
OY	421 KLLASOKCOLATTISPISHSHOEADULLROAKKLPCEPIIRRAYADIATINLTMDPEKK	480			
Db	419 RILTSQKPLAKAAIAELUSTVAHPQALSVSKAALTPEDPILIRAVANIALVTMTODPEKK	478			
OY	481 RSLDYVAKKLLQEOFLLEFDTERNOPRHSMPYLRYQVTPESRKTMLMLDLLETVALSKSED	540			

Db 479 ALXQVAEQLIGDIIIFTEDEBNPLPSHSSYLKQVSPETRSQLMIIETLIVSSKTDED 538

QY 541 IRLLIQIMTEBGDAKNPVPVLAGLIIKIVE 568

Db 539 IRVFLSLMKKTHYKNIPILSSGLMRIYE 566

RESULT	4
G71525	

Probable outer membrane leader peptide (omp) CT350 - Chlamydia trachomatis (serotype D, C); Species: Chlamydia trachomatis
C; Date: 13-Sep-1998 #revision 13-Sep-1998 #text_change 04-Feb-2000
C; Accession: G71525
R; Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A; Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis; Reference number: A71570; MUID:99000809
A; Accession: G71525
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-566 <ANN>
A; Cross-references: GB:AE001308; GB:AE001273; NID:J3338766; PIDN:AAC67945.1; PID:J3338767
A; Experimental source: serotype D, strain UW-3/Cx
C; Genetics:
A; Gene: CT350

Query Match 50.2%, Score 1421, DB 2; Length 566;
Best Local Similarity 53.0%, Pred. NO. 1.5e-78;
Matches 301; Conservative 88; Mismatches 177; Indels 2; Gaps 1.

A:12043
 hypothetical protein alr1903 [imported] - Anabaena sp. (strain PCC 7120)
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
 A:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
 C:Accession: A12043
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Iriki,
 N.; Nakazaki, N.; Shikpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: A12043
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1547 <KUR>
 A:Cross-references: GB:BA000019; PUDN:BA073602.1; PID:g17130993; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 C:Gene: alr1903

QueryMatch	7.2%	Score 203;	DB 2;	length 1547;
Best Local Similarity	21.3%	Pred. No.	0.00037;	
Matches 114;	Conservative 105;	Mismatches 239;	Indels 76;	Gaps 16;

RESULT 6

AB2239
hypothetical protein all3465 (Imported) - *Anabaena* sp. (strain PCC 7120)
C.Species: *Anabaena* sp. (strain PCC 7120) is a synonym of *Nostoc* sp. strain PCC 7120
A.Note: *Anabaena* sp. (strain PCC 7120) is a synonym of *Nostoc* sp. strain PCC 7120
C.Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C.Accession: AB2239
R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irida,
N.; Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A.Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

[illegible]

Query Match 5.1%; Score 145; DB 2; Length 1010;
Best Local Similarity 22.3%; Pred. No. 0.66; Mismatches 136; Conservative 84; Mismatches 210; Indels 180; Gaps 30;

Query 18 ISLVAKPEFSGVGHKILYISTQTOALATYVEALDAV--GDHDFVLKRIGEDYLKOSIH 75
Db 81 IALADKLPPDL-----LSEALTSAREIQDEYICADALIALAKRLPPDLSEALA 129
QY 76 SS----DPQTRKSTITIGAGLASSFALDYLSQAMETADPQQQLVLSVSGHLGKTSDDL 131
Db 130 TAREIQDEYFRTSTLIEI--AEKLPVLSFALAAAREIQ-----DE 168
QY 132 LFK-----ALSPYVILDEAAVRLANKNKRV-IDHHSFIHKLPEEI--QCLSAATFL 183
Db 169 YFRASTLIALAEKLPVLSFALAAAREIQDEYFRADARELAQKLPDLSEALAAAREI 228
QY 184 RLETESDAVIRDLAAKSAIRSATAIQIGEQKRF-LPTRLMLTSSAFQDEALY 242
Db 229 QPEYIRADALI--ALVEKLPVLSFALAAAREIQDEYIHAADRLRYOKLPDDLGEVIA 286
QY 243 ALGKIDGQSYNNIKQLOK--PD-VDVTLAAO-----ALIALGKE-----E 282
Db 287 AATEIRGCVPHNPLRELAELPDLSEALAAAREIODESNSRAHALRELAELKLPDLS 346
QY 283 DALPIYK-----QALEER-----PRAIYALRHLPSEIG----- 311
Db 347 EALTATREIOSEHYHASTLRLAOKLPDLSEALAAAREIODESNSRASTLELAELKPS 406
QY 312 -IPALPFELTKNSEAKLNVALLE-LGCTPKILEYTERVQPHYNETLALFSKG 369
Db 407 VLPELAAVARKIRKHSNAYGILALAEKLPVLPALAAAT--LEPEYHRA----- 456
QY 370 RTLQNMKRVNIIVDPQDERERLLSTTRGLEEQILTFPLRPKEAYLPCYKLLASQTKQ 429
Db 457 -----STURELAE-----KLPPD-----LLSE--- 473
QY 430 LATTAISFLSHTSQHEALDLFQAAKLPGEPIIRAYADLAITNLTKDPKKRSLNDYAKK 489
Db 474 -ALTAISEIQPKSNR-ADALIALAEKLPDLSEALA--AIREIODESNSRAHALALAEK 529
QY 490 ---LIOETLLE--VDTENORPHSPMYLRQVPPESRTKMLDLI-ETLATSK-----S 537
Db 530 LPPDLSEALAAIREIQODESNSRAHALIALAQ-----KLPPDLSEALAAATREIQSK 580
QY 538 SEDIRLLIQL 547
Db 581 SNRVHALLAL 590

RESULT 10
AF2010 regulatory protein all1636 [imported] - Anabaena sp. (strain PCC 7120)
C.Species: Anabaena sp.
A.Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C.Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C.Accession: AF2010
R.Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri-
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata
DNA Res. 8, 205-213, 2001
A.Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A.Reference numbers: AB1807; M01D:21595285; PMID:11759840
A.Accession: AF2010
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1381 <KUR>
A.Cross-references: GB:BA000019; PIDN:BA078002.1; PID:917135456; GSPDB:GN00179
C.Experimental source: strain PCC 7120
C.Genetics:
A:Gene: all1636

Matches 122; Conservative 87; Mismatches 190; Indels 190; Gaps 23;

QY 20 LVAKPEESGHIKITYISTOSTOQALATYIEALDAYGDHPFV--LAKIGEDYIKOSIHSS 77
 Db 688 LVDKLPELI-----PEALAAAREIENSYRRDALISLANKLPPELLPEAVVAAA 735

QY 78 ---DPQTRKSTIGA-----GLASSEALDVLGSA---METADPLQOOL----- 115
 Db 736 REIOEANNAEVLGSLDKLPPELLPEALAAAREIEFESSRATYILSLAKLPPELLPEALA 795

QY 116 -----VLANSVGH-----GKTSDDLKFA-----LASPPYVIRLE 146
 Db 796 ATREIONGYNRADILSLADKLPPELLSEALAVAREIDEKRYAREALSLADKLPPELLPE 855

QY 147 AAYRLANKNTKV-IDHHSFTHKLPEIQCCSAIFLR-LTESDADATIRLLAAK--- 201
 Db 856 ALAAAREIDEKRYNDALSLANKLPPELL--LSEAVVAAREIDEKRYRAATYILSLAAKPE 913

QY 202 -----KSAIRSATAIOLIGEYQOKRFLPLRLN 228
 Db 914 LPEALAAAREIOFESSRAQALRSLADKLPPELLPEALAAAREIODEBYRAQ-----ALRSL 969

QY 229 LFSAPQODEALLYLAKIKDQOSTYNNIKQL--QKPD--VDVTLAAQALALGKEEDAL 285
 Db 970 ADKLPEPELLPEVVAATREIEYGEFRASILGSLAVQLPELLPEAVAAAREIYDAFHSDAL 1029

QY 286 PVKKOALEERPRLAYALRH-----LPSELGIFLALPIFLKTKNSEAK 328
 Db 1030 SMLAKKFEVLPPEALAAAREIIEHCRADSSSLADKLPPELL--LPEAVAAAREIDEKRYR 1088

QY 339 LNVALLLELGGDPKLL--EYITERLYOPHYNETLALSFSGKRTLOWMKRVNITVPODP 386
 Db 1089 ---AVVLISLADKFPPELLPEAVAAARGIODEYPRSLAL-----RLAD-----KLIP 1131

QY 387 QERELISTTGLLEQOILTFLEPLKPEAYLPCTYKTLASOKTQALATYATISFLSHTSHOEA 446
 Db 1132 ELPEALAAAREIODD-----GHMRA 1152

QY 447 LDILFQAKLPGEPILIRAYADLAIYNLTKDPEKKRSLHDYAKKLIQETL 495
 Db 1153 YNLSDLAKLPPELLPKALA--AAAREIODEYRRANALSLADKLPPELL 1199

RESULT 11
 S28294
 hypothetical protein ZC84.3a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 12-Mar-1993 #sequence, revision 12-Mar-1993 #text, change 30-Sep-1993
 C:Accession: S28294
 R:Thomas, K.
 submitted to the EMBL Data Library, December 1992
 A:Reference number: S28285
 A:Accession: S28294
 A:Molecule type: DNA
 A:Residues: 1-730 <THO>
 A:Cross-references: EMBL:219157
 C:Genetics:
 A:Insertions: 77/3; 120/3; 213/1; 228/2; 254/3; 332/3; 428/2; 499/3; 687/3; 727/3

Query Match 4.7%; Score 134.5; DB 2; Length 730;
 Best Local Similarity 20.5%; Pred. No. 1.8;
 Matches 139; Conservative 101; Mismatches 250; Indels 187; Gaps 27;

QY 1 MGLFHLTLFGL-----LLGSLPISLVAKFPESVGHKITYISTOSTOQALATYIEAL 51
 Db 9 LGCEELAVKDLRSQVLRREAAITCSFIVSKYGIETHSIGEDILV--PAMSQVAAVSTKIMAT 66

QY 52 DAYGDHDFEVLARKIGEDYLKOSIHSSDPQTRKSTIIIGAGLAGSEALDVLGSAQMETADPL 111
 Db 67 SASLTTFEIV-----EYV-----QTR-----QVFTIILSFSTSKDKS 98

QY 112 QQLVLIVSAVSGHLGKTSDLLFKALASPYPVIR-----LEAAVRLANLKNKTVIDHLSF 166

Db 99 QRQLALALLEIYTSKWSDRNKQIMROICELIKSAINDADSETRAAGRAFAKIDBMHS- 157
 QY 167 IHKLPEEIOCL-----SAAIFLRETESDADATIRLLAAKSAISASATA 210
 Db 158 -----EADALVLELDHOKMKLRGDDAASNAVSASEKSTIPIRKSLGASGAHNNISA 212

QY 211 LOIGEQOKRFLPTLRNLLTSASPQODEALIVLALGKDGOS-----YYNIRKQLOKPD 264
 Db 213 ISEKKEGIKKLPIYAD--TSLNPIEIKNIGCNLRLSDASNTWVLEYISFVRTHSR 270

QY 265 VDVTLAAQALALGKEEDALPVKKOALEERPRLAYALRHLPSELGIFLALPIFLKTKN 324
 Db 271 LSEWLRLALAKLPARKAAETLPLTKKQ-----IGHTLVNILECPN 310

QY 325 SEAKLNVALLALELGD-----TPK-----LLEYITERLYOPHYNETLALSFSG-----R 370
 Db 311 ANHQL--VTVELMCDPRLHMLVPRARVVLLEYLISL--DEYTEGASINAKELKTAIR 365

QY 371 TLQNMKRVNITVPODPQERELISTTGLLEQOILTFLEPLKPEAYLPCTYKTLASOK--- 427
 Db 366 KMLTW-----ASDP--RLSILTPH--VERAKISMFCVNVADFSALISDIDSEQKNMI 414

QY 428 -----TOLATTAISFLSHTSHOE--ALDILFQAKLPGEPI----- 461
 Db 415 HOTLQRLNGLENGISSNNIATNSGATASRSTSTQKESTSGLPEFGARKGCTGVNIGS 474

QY 462 IRAYADLAIYNL-----TKDP-----EKKRLHDYAKKLIQETLIFYDT 500
 Db 475 LINSNNLALSRLEQSTSRMEKVMNLSTVLPPTLRIQVQD-----LQKMSSSEA 530

QY 501 ENQRHPSMPYLR-----YQVTPESRTKLMDILELATSKSESD---IRLLQLMTEG 551
 Db 531 DEQESAISSIYMWICDGGFGEWCQYAKLLNLFELLSRSBENKMKCLRILIGKMTAQ 590

QY 552 DAKNPPVLAGLIRIKIVE 568
 Db 591 AAKLPDSTEMAVCKYLD 607

RESULT 12
 C88550
 protein ZC84.3 (Imported) - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence, revision 10-May-2001 #text, change 10-May-2001
 C:Accession: C88550
 R:anonymous, The C. elegans Sequencing Consortium.
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
 Science 282, 2012-2018, 1998
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C-
 A:Accession: C88550
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-826 <STO>
 A:Cross-references: GB:chr_III; PIDN:CA79568.1; PID:g3881445; GSPDB:GN00021; CESP:ZC
 C:Genetics:
 A:Gene: ZC84.3
 A:Map position: 3

Query Match 4.7%; Score 134.5; DB 2; Length 826;
 Best Local Similarity 20.5%; Pred. No. 2.1;
 Matches 139; Conservative 101; Mismatches 250; Indels 187; Gaps 27;

QY 1 MGLFHLTLFGL-----LLGSLPISLVAKFPESVGHKITYISTOSTOQALATYIEAL 51
 Db 89 LGCEELAVKDLRSQVLRREAAITCSFIVSKYGIETHSIGEDILV--PAMSQVAAVSTKIMAT 146

QY 52 DAYGDHDFEVLARKIGEDYLKOSIHSSDPQTRKSTIIIGAGLAGSEALDVLGSAQMETADPL 111
 Db 147 SASLTTFEIV-----EYV-----QTR-----QVFTIILSFSTSKDKS 178

```

QY 112 QQLLVLSAVSGHLGKTSDDLFFKALSPYVIR-----LEAAVRLANIKNTKVIDHLASF 166
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 QRRQLALLETIVISKMSDRKKQIMRQICELISAIMDADSEPPRAAGRRPAKIDENHS- 237
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 IHKLPEETIOQL-----SAAIFLRLTEESDAYTRDLAAKSAIRSATA 210
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 -----EADALVLELDHDSKOKMLRGDPAASSVNSEKSIPIRSKLSGAKAHMNTA 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 211 LQIGEVQQRKFLPTLRNLKLSASPODEALIVLNGRIKADQS-----YNIKKQLOKPD 264
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 ISEKKEGIRKLLPVAD--TSLNPTETKNGCNLRLSDASNTMLEIYSIFVRIHSSR 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 VDVTLAAQAQLIALGKEDALPVYKQALEERPRALVALRHLPSEIGIPALPIFLTKN 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 LSEMLRLALAKLFARKAAETLPNTKKQ-----IGHLVNILECFN 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 SEAKLVNALVLEIGCD-----TPK-----LLEYITERLVOPHYNETLALSPSKG-----R 370
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 AHHQI---VTVCEIMCDPIHLMVPRKARVLELYTSL--DEYTEPGASINAKELTAIR 445
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 TLQNMKRVNIIVQDPERERLLSTTRGLEOILTFRLPKKAYLPCITKLLASQK--- 427
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 446 KMLTW-----ASDP--RLSILLTPH--VEKALCSMFQYVADFSALISDLSEQKNWI 494
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 428 -----TQLATTAISFLSHTSQK---ALDILFQAALPGEPPI---- 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 495 HQTLQRRGLENGISSNNIAINSGATASRETSNNTSFOKESTFGLPECGAKGGGVMLGS 554
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 462 IRAYADIAIYNL-----TKDP---EKRSKLDHVAKKLIOETLLFVDI 500
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 LINSNNILRLSEBOSTSRLEMEKYNLSTVLPDPDTLEKIONVQD---LLQKRSENA 610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 501 ENQRRPHSMRYLR-----YQVTPESRKMLDILETLATSKSSD---IRLLIQMTREG 551
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 611 DEQSAISSTYMMICDGFVEWCYAKLLNLFELLSKRSSENNKRCILRIIGKMQTAQ 670
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 552 DAKNFVLAGLILIKIVE 568
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 AAKLFDETEMAVCVKVD 687
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
S17944
virulence sensor protein bygs precursor - Bordetella bronchiseptica
C:Species: Bordetella bronchiseptica
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S17944
R:Atico, B.; Scarlato, V.; Monack, D.M.; Falkow, S.; Rappuoli, R.
Mol. Microbiol. 5, 2481-2491, 1991
A:Title: Structural and genetic analysis of the byg locus in Bordetella species.
A:Reference number: S17943; MUID:92167813
A:Accession: S17944
A:Molecule type: DNA
A:Residues: 1-1238 <ARI>
A:Cross-references: EMBL:X58355; NID:g39352; PIDN:CAA41252.1; PID:g39354
A:Experimental source: strain 7865
A:Note: the authors translated the codon GAG for residue 134 as Gly and TCC for residue
C:Function:
A:Description: involved in signal transduction
A:Superfamily: evgs protein; response regulator homology
C:Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase;
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1238/Product: virulence sensor protein bygs #status predicted <MAT>
F:698-945/Domain: sensor histidine kinase homology <SHK>
F:975-1091/Domain: response regulator homology <RRH>
F:729/Active site: His (phosphohistidine intermediate) #status predicted
F:1023/Binding site: phosphate (Asp) (covalent) #status predicted

```

```

QY 19 SLVAKPEESVGHKILYISTOSTQOALTYLEALDVGDDHFFVLRKIGEDYLKOSIHSD 78
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 TLQQAYVQATLHYF-----PSSBQALAAV-----AYQADVFIGDALTTSHVLSQSIFND 232
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 PQT-----RKSTIIGAGLAGSSEALDVL 101
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 VRVAPAQIVTGESFQVRADNTRLRVNNAVLEAIPASERRSRLTYRMGL--GSSISIDFA 291
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 102 SQAMETAD-----PLQQLLVLSAVSGHLGKTSDDLFFKALSPYVIRLE----- 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 RPAYSAREQOMMANHPVAKVAVLN-----LEAPFTLFTDEQFGGISA 334
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 AAVRYLANLK---NTRKVI--DHLHSFIHKLPREIQLTSAIFRLRETEE---SDAYIRD- 196
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 335 AVIQLDLQRLGDLQIIGVDTVEELIKLSGEDMAGALFVNAAARSVLSFSRPYVRNG 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 -LLAAKSAIRSATAVLOIGEVQQRKFLPTLRN-----LTSASPODQ-----EAILY 242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 LVIVTRQDPAAPADA---DHLQDRTIAMVRNSAIFLLQQRYPQAKRVVADNPTeamLL 450
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 ALGKLKQKQ-----SYT-----NIKKQLOKPDVUTVIAA----- 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 ---VADGQADAVVQTOISASYVNRYPAGKLRITASALDPPAPETALATAGOTELISIL 506
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 273 -QALIALGKEDALPVYKQALEERPRALVALR---HLPSRIGIPILPIFL----- 320
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 NKALYSISNDELASIVSRKWSGDDPRTWYAIRNEYILLIGLISL-LFLSIVYVLR 565
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 321 ---RTKNSKALVNALVLEIGCD--TPKLEYITER-----LVOPHYNETLALSPSK--G 369
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 QIRQKRAERANLQLEFRMYRLIDGTFNPI-YVADKGGRMCLNDAYLDIFGVADAVALG 624
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 370 RLQNMKRVNIIVQDPERERLLSTTRGLEOILTFRLPKKAYLPCITKLLASQKQ 429
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 625 KTI---PEANVY--GDP-----ALAREMEHFLTT-----RMAERPR 657
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 430 LATTAFSLSHTSK--OEAIDLFLQAALPGEPPIRAYADIAIYNLTKDPKKKSLHD-- 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 FEDRDVTLHGKTRHYVQWTVYVYSGSLGKQ--IIGGMID-----ITERAELLRELDHAK 710
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 -----YAKKLIQET-----LLFVDENORPHSPMYLRVQVTPESR 521
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 ESADAARAKTTFLATWSHEIRTPMNAIIGMLELALLRPADQEPDRQSIGVAYD-SARSL 769
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 522 TKMLDILETLATSKSSD---IRLLIQMTREGAKNPVLA 560
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 770 LELIGDILIDIAKIEAGREKFDLAPVPTALRALPEGAIRVFDGLA 811
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
S17946
virulence sensor protein bygs precursor - Bordetella parapertussis
C:Species: Bordetella parapertussis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: S17946
R:Atico, B.; Scarlato, V.; Monack, D.M.; Falkow, S.; Rappuoli, R.
Mol. Microbiol. 5, 2481-2491, 1991
A:Title: Structural and genetic analysis of the byg locus in Bordetella species.
A:Reference number: S17943; MUID:92167813
A:Accession: S17946
A:Molecule type: DNA
A:Residues: 1-1238 <ARI>
A:Cross-references: EMBL:X52948; NID:g39727; PIDN:CAA37124.1; PID:g39729
A:Experimental source: strain 9305
A:Note: the authors translated the codon GAG for residue 134 as Gly
C:Function:
A:Description: involved in signal transduction
A:Superfamily: evgs protein; response regulator homology
C:Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase;
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1238/Product: virulence sensor protein bygs #status predicted <MAT>
F:698-945/Domain: sensor histidine kinase homology <SHK>

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 01:14:10 : Search time 2366.15 Seconds
(Without alignments)
16865.755 Million cell updates/sec

Title: US-09-662-812-1

Perfect score: 1907
Sequence: 1 gtgctgtgatttgaaaaaag.....ctcatcaagagcgttgaa 1907

Scoring table:
OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 segs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	1907	100.0	1907	6	AX100528	AX100528 Sequence
2	1907	100.0	8268	1	AE002234	AE002234 Chlamydia
3	1907	100.0	16448	1	AE001587	AE001587 Chlamydia
4	1707	89.5	1707	6	AX349527	AX349527 Sequence
5	84	4.4	2088	6	AX349569	AX349569 Sequence
6	24	1.3	36	6	AX100531	AX100531 Sequence
7	24	1.3	44	6	AX100530	AX100530 Sequence
8	22	1.2	11627	1	AE001308	AE001308 Chlamydia
9	22	1.2	47445	2	AC017438	AC017438 Drosophila
10	22	1.2	183722	3	AC009341	AC009341 Drosophila
11	22	1.2	261000	3	AE003654	AE003654 Drosophila
12	21	1.1	4343	6	AX035433	AX035433 Sequence
13	21	1.1	177014	2	AL670958	AL670958 Mus muscu
14	21	1.1	206591	2	AL611984	AL611984 Mus muscu
15	21	1.1	225609	2	AL645545	AL645545 Mus muscu
16	20	1.0	558	10	MUSRABAP	MUSRABAP Mus muscu
17	20	1.0	2680	5	AMWNTIR	AMWNTIR Mus muscu
18	20	1.0	3330	8	AF024648	AF024648 Arabidops
19	20	1.0	3620	10	BC022119	BC022119 Mus muscu
20	20	1.0	4907	10	MMU245569	MMU245569 Mus muscu
21	20	1.0	17854	1	AF307053	AF307053 Thermococ
22	20	1.0	18001	1	AF307052	AF307052 Pyrococcu
23	20	1.0	23619	2	AC102470	AC102470 Mus muscu
24	20	1.0	35857	2	AC026146	AC026146 Homo sapi
25	20	1.0	52347	2	AC068214	AC068214 Homo sapi
26	20	1.0	68996	2	AC102790	AC102790 Homo sapi
27	20	1.0	86212	2	AC106082	AC106082 Rattus no
28	20	1.0	104534	9	AL356317	AL356317 Human DNA
29	20	1.0	104944	9	AL591509	AL591509 Human DNA
30	20	1.0	105168	9	AC091317	AC091317 Homo sapi
31	20	1.0	114567	9	AC010473	AC010473 Homo sapi
32	20	1.0	124146	9	AC068706	AC068706 Homo sapi
33	20	1.0	131691	9	AC096765	AC096765 Homo sapi
34	20	1.0	139616	2	AL445261	AL445261 Homo sapi
35	20	1.0	146935	9	AC096647	AC096647 Homo sapi
36	20	1.0	148845	9	HS114A1	HS114A1 Human DNA
37	20	1.0	149671	9	HS376D21	HS376D21 Human DNA
38	20	1.0	151976	9	AC024588	AC024588 Homo sapi
39	20	1.0	156785	2	AC018541	AC018541 Homo sapi
40	20	1.0	164394	2	AC022873	AC022873 Homo sapi
41	20	1.0	164819	9	AC092335	AC092335 Homo sapi
42	20	1.0	166692	2	AC023475	AC023475 Homo sapi
43	20	1.0	167332	2	AC080168	AC080168 Mus muscu
44	20	1.0	170736	2	AC013658	AC013658 Homo sapi
45	20	1.0	170933	2	AC087619	AC087619 Homo sapi

ALIGNMENTS

RESULT 1
AX100528
LOCUS AX100528 1907 bp DNA linear PAT 10-APR-2001
DEFINITION Sequence 1 from Patent WO0121804.
ACCESSION AX100528
VERSION AX100528.1 GI:13619532

SOURCE
ORGANISM Chlamydia pneumoniae.
Chlamydia pneumoniae
Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales
REFERENCE
Murdin, A.D., Oomen, R.P., Wang, J. and Dunn, P.
I Chlamydia /i antigens and corresponding dna fragments and uses
thereof
Patent: WO 0121804-A 1 29-MAR-2001;
Aventis Pasteur MSD (FR)
Location/Qualifiers
1..1907

FEATURES
source
1..1907

CDs
/organism="Chlamydia pneumoniae"
/db_xref="taxon:83558"
101..1807
/note="unnamed protein product"
/codon_start=1

```
/transl_table=1
/db_xref="GI:13619533"
/transl_table="MGLFHLIFGLLCSLIPISLVAKPESVGHKIIYISTOSTOAL
ATYLEADAYGDHDFVLRKIGEDYLKOSIHSDPQKSTIISAGLAGSEADVLIS
QAMETADPLIQLLVLSASGHILKSDLEKALASPVPKRLSEAYLAKMKV
DHLSEIHKLPDEIOLSAATFELRESESAVIRDLAAKSAIRSATLOIEYOO
KRRLPTLRILTSASPODEAIIYALSKLKGCOSYINIKLOKRPDVTILAAOALI
ALKEEDALPIYKQALBERRALYALHNPSEIGIPALFIFIKTNSEKLVALLA
LLEIGDTPKLEYITERLVOPHNELTSLFSGRILQNMKRVNIIVPODERERL
LSTRGEOELTFELFRLPEKAYLPCYIKLASOKTQIATAISLSTSHOEDL
FOAKLPGEPILIRAYADLAIVNLKDEKSLDHYAKKLIQETLLEFVDTENOPHS
MPYRYQVTPESRTRKMLDIETLATSSSDIRLLIQLMEDGAKNPVLAGLLIKI
VE"
BASE COUNT      542 a      456 c      380 g      529 t
ORIGIN
Query Match      100.0%; Score 1907; DB 6; Length 1907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtgagctatttgaaaaagtcacatgagatggttataatgltcaaggtctccatcc 60
Db 1 GTGCGTGTATTTGAAAAAGTCATGATGTTTATATGTTCAAGGTCTCCCTATCC 60
QY 61 aaactgaactactgctagagaggttgaaacatcgatctatgagactatccatcaac 120
Db 61 AAACATTAATAACTTGTGTAGAGGAGTGAACATGATCTATGAGCATTCATCTAAC 120
QY 121 tctcttgagacttattggttagctctccatctctctctctgttgtaattcccgagtc 180
Db 121 TCTCTTGAGACTTTATGTGTAGCTCTCCATTTCTCTTGTGCTTAATTCCTCGAGTC 180
QY 181 tgaagtcataagatcccttatalaaglacgcaatctacacagcagcgcttagacaata 240
Db 181 TGTAGGTATTAAGATCTTTATATATAGTACGCAATCTACACAGCGGCTTAGCAACATA 240
QY 241 tctggaagctctagatgctcagatgatacagactctctctgtttaaagaataatcgaga 300
Db 241 TCTGGAAAGCTCTAGATGCTTACGGGTGATGATCATCTTCTGTTTAAAGAAAAATCGAGAGA 300
QY 301 agactatctcaagcaaaagcatccactctcagatccgcaaaatagaanaagcacatcat 360
Db 301 AGACTATCTCAAGCAAAAGCATCCACTCTCGATCCGCAAACTAGAAAAAGACCATCAT 360
QY 361 tggagcagcgctggcgagatcttcagaaagccttggagcgtgtctcccaagctatgaaac 420
Db 361 TGGAGCAGCGCTGGCGGAGCTTTCAGAAAGCCTTGAGCGTGTCTCCCAAGCATGAGAAAC 420
QY 421 tgcagagccccctgcagcagcagctggtttatcgcgagctcagagacatcttgggaaac 480
Db 421 TGCAGAGCCCCCTGCAGCAGCAGCTGCTTTATCGGCACTTCAGAGACATCTTGGGAAAAAC 480
QY 481 ttctgaagcactactggtttaagcttttagacatcccatctcgttcacacgcttagaagc 540
Db 481 TTCTGAGCAGCTTACTGTTTAAAGCTTTTAGCATTCCTCATCTGTCATCCGCTTAGAAGC 540
QY 541 cgactatagacttgctaaattgaaagaacataagtcaltgacatctacatcttcttcat 600
Db 541 CGCCTATAGACTTGCTATTTGAAGAACACTAAAGTCAATCATCATCTACATCTTCTTCAT 600
QY 601 tcaataagctccgaaagaataccaatgctcatctcgcgcaaatatctcctagccttggagac 660
Db 601 TCATAAAGCTTCCGAAAGAAATCCAAATGCTATCTGCGGCAATATTTCCATCCCTTGAGAAC 660
QY 661 tgaagaatctgagtctatatctcgagatctcttagctgcaagaaagcgagatctcgag 720
Db 661 TGAAGAATCTGATCTTATATTTGCGGATCTCTTAGCTGCGCAAGAAAGCGGATTCGAG 720
QY 721 tgcacagccttgcagatcggaagataccaaacaaacgcttcttcgcagactttagaa 780
Db 721 TGCACAGCCTTGCAGATCGGAGATACCAACAAACAAAGCGTCTTCTTCGACACTTAGGAA 780
```

```
QY 781 ttctgtaacgagtcgctctcctcaagatcaagaactatcttcttgctttagggagct 840
Db 781 TTTGCTAACGAGTGGCTGCTCCCAAGATCAACAAAGCATATCTTATGCTTTAGGCAAGCT 840
QY 841 taagatgtgcaagctcctacatatataaagaatctgcaagacccgagatgtagt 900
Db 841 TAAGATGTGTGAGCTCTACTACATATATAAAGCAATTTGCAAGAGCTGATGTGATGT 900
QY 901 cactttagcagcagctcactcaagctttaaattgcttgggaaagaagagcgcctcccgct 960
Db 901 CACTTTAGAGAGAGCTTCAAGCTTTAATGTCTTTGGGGAAGAGAGAGAGCGCTCTCCCGT 960
QY 961 gataaaaaagcagcacttgagagcgcgctcgagcctgtatgctccttaagcactac 1020
Db 961 GATFAAAAAAGCAAGCACTTGAAGAGCGCGCTCGACCCCTGTATGCTTACGCACTATCC 1020
QY 1021 cctcagatagagatcccgatcgccctcgatctccgaatctccaaataaagaagaagcagc 1080
Db 1021 CTTCTAGATAGGAGATTCGATTCGATTCGCTCGCATATTTCTTAAACATTAAGAACGCGAAC 1080
QY 1081 caagltgagtagctttagctctctcttagagtgtaggtgtagcacccctaaactactgga 1140
Db 1081 CAAGTTGATGTAGCTTTAGCTCTCTTAGAGTTAGGTGAGGAGACCCCTAAACATCTGGA 1140
QY 1141 atacatcaccgaaggtctgtccaaacacatataatgaagactcagccttgaattctc 1200
Db 1141 ATACATTTACCGAAAGCTTGTCCAAACACATTTATATAGACTTACCTTGAAGTTTCTC 1200
QY 1201 taaaggcgctacttacaanaattggaagcggttgaaacatcaatgacccctcaagatcccca 1260
Db 1201 TAAGGCGGCTACTTTACAAATTTGGAAGCGGCTGAAACATCATATGCTCTCAAGATCCCA 1260
QY 1261 ggaagggaagaggtgtcctccacacacagaggtcttgaagagcagatccttact 1320
Db 1261 GGAGAGGGAAGAGGTGTCTCTCCACAAACCGAGGCTTGAAGAGCAGATCTTACTTCT 1320
QY 1321 ctccgctacactaaagaagcttacctccctgtattataagcttctggcgagtcagaa 1380
Db 1321 CTTCCGCTACTTAAAGAAAGCTTACCTCCCTGTATTTATAGCTTTGGGAGTCAAGAA 1380
QY 1381 aactcagcttgcaactactgagatctcttctttaaagtaacacctacaatcaggaagcctt 1440
Db 1381 AACTCAGCTTGCACTACTGCGATTTCTTTTATAGTCAACCTCACATCAGAGAGCCTT 1440
QY 1441 agatctacttctcaagcttgcgaagcttcccttggagaacctatcaltcgcgcgtatgcaga 1500
Db 1441 AGATCTACTTTTCCAAAGCTGGAGAGCTTCTGGAGAACCTATCATCCGCGCTATGCAGA 1500
QY 1501 tcttgctattataatctccaccaaagatccttgaaaaaaagcttctccatgatattgc 1560
Db 1501 TCTTGCTATTTATATATCTCACCAAGATCTGAAAAAAGGTTCTTCCATGATATTATGC 1560
QY 1561 aaaaaagctaatcagaagaacgttgaatttggtagacggaacaaacaaagacccatcc 1620
Db 1561 AAAAAAGCTAATTCAGGAACCTGTTTATTTGTGACACGGAACCAACCAAGCCCATCC 1620
QY 1621 cagatagccctatactagcttlaacaggtcaccccaagaagcgtagaagctcatgltgga 1680
Db 1621 CAGCATGCCATCTAGCTTATCAGGTTCACGCCCAAGAAAGCGGTAGAGAGCTCATGTTGGA 1680
QY 1681 tatcttagagacactagcaccctcgaagctctccgaagatactcggttatttgatacaact 1740
Db 1681 TATCTTAGAGACACTAGCCACCTGGAAGTCTTCGGAAGATATCTCGTTATATGATACACT 1740
QY 1741 gatagaggaaggaatgcaaaaaatctcccgagctccttgcaagcttaccataaanaatgt 1800
Db 1741 GATGACGAAGGAAGATGCAAAAAATTTCTCCAGTCTTGCAGGCTTACTCATATAAATTTGT 1800
QY 1801 ggaataaccccaactacgtcttlaagaaagcttgccttcttcttagactcctctgtctc 1860
Db 1801 GGAATTAACCCCAACTACGCTTATGAAAGCTTGCCTTCTTATTTCTAGCTTCTTGTTC 1860
```

Qy 1861 ttatgggttcctcaagctatgcttctgcatcatcaagagctgtgaaa 1907
|||||
Db 1861 TTTATGGGTTCTCAGCGATGCTTTTGACTCATCAAGAGGCTGTGAAA 1907
|||||

RESULT 2
AE002234 8268 bp DNA linear BCT 30-MAY-2000
LOCUS Chlamydomophila pneumoniae AR39, section 62 of 94 of the complete
DEFINITION genome.
ACCESSION AE002234 AE002161
VERSION AE002234.2 GI:8163492
KEYWORDS
SOURCE Chlamydomophila pneumoniae AR39.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE 1 (bases 1 to 8268)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNAL MEDLINE
PUBMED 20150255
10684935
REFERENCES 2 (bases 1 to 8268)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7189667.
LOCATION/Qualifiers
1. 8268
/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
113. .1036
/gene="CP0751"
113. .1036
/note="similar to GB:U00096 SP:Q47012 PID:1381661
PID:1788603; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="AtsA/FlaC family protein"
/protein_id="AF38556.1"
/db_xref="GI:7189668"
/translation="MSSRELIIIGSSQOQPTRTNQGAVLFNNGEGLFDPEEGTOR
QTFANAPATVNRIVFVSHFGDHCGLGSLMRNLNDVSHPIHCYVYASGKYDR
LRGTGYHETIIOVHEHPISEEGVEDSEFTEAORLOHOVDLIGRITPEPTIKLP
KLEESRGILTIODLIRPOEISIGSSTYVLSVSVYRKGDSTAIITADLLPGCAALDD
AKNSCMLCESSTYLEDQHRHLAESHFTMTAKQATLAKRATQKLLITLHFSARYLNDL
FKKEASVFPNPVSAOERYSTPEPKNPLNK"
1121. .2059
/gene="CP0752"
1121. .2059
/gene="CP0752"
/note="Integrase/recombinase, phage integrase family;
identified by match to Pfam protein family HMM PF00589"
/codon_start=1
/transl_table=11
/product="Integrase/recombinase, phage integrase family"
/protein_id="AF73701.1"
/db_xref="GI:8163493"
/translation="MTASIVSFLDYLKMWKSAPHTLRNYCLDNLNGLKIFLEERGNIA
PSSPIQLATEKRVSELPFSLFTEKHEVRMYIAKLINENKAKRTIKRCLSSIKSFAHYC

gene
CDS
2114. .3706
/gene="CP0753"
2114. .3706
/gene="CP0753"
/note="ABC transporter, ATP-binding protein; identified by
match to Pfam protein family HMM PF00005"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AF73702.1"
/db_xref="GI:8163494"
/translation="MSTVLDKICKSLGTRILPDVSVENPNCYGLTGPNGAKSTL
LKTIKMGIEPTRGISILPKKVGILRQNIIDSFHTVLDVCYIKMSTRIMEALQRRNLY
LQFTDAIGMEIGEIEIIGENGVRADSEABELIGIGIPNMPDKKAMIPIDLOF
RYLCOALGHPHALLIDEPNTNHLDSYINMGNFLKDEGYIVVSHDHPNLTIT
HIADIDYITIIYPGNVDDMVEMKTSRROEKADIKSEKKSISOLKEPVAKPGAGSRA
SOVOSRLREIKIQPOEIKSNRORXYIRPISDSKSGYVLSLAETIDYGDHOVYH
PSLEIYQGDKLGITGNNGIGKTLTKLAGVAPBSGSIKLGHOAICSTFPONHSDV
LADCGGETLFEWLRNRKTGINDEIRSVLGKMLGDDDAFKQIALSGGETARLLNAG
MLENHNVLILDEANNHLLDESYSALSMAINDYKTAIFVSHDRGLIDCATKLLIFD
KDKITFPFGTMVDYTAGHKOLL"
3735. .4325
/gene="CP0754"
3735. .4325
/gene="CP0754"
/note="similar to GB:L08793 SP:Q02169 GB:M96343 PID:142853
PID:143161; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="mat protein"
/protein_id="AF38557.1"
/db_xref="GI:7189669"
/translation="MSLPVLGSSPRRKFIIEKFPVFTVIPSNFDEKSYSGDPI
AYNGELAAKAVAVSELHSPDCIITLGDYIVDGRIFEPQDKADAIQMLKLNQ
THDVYSTAVYIHKGKLTGSETSQISLTMPPIRISYIDYVGTLLNCGAIVDCRGL
IKKHVGCYINVOGLPIQTLKYLEELNDLMDYST"
4307. .6013
/gene="CP0755"
4307. .6013
/gene="CP0755"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AF38558.1"
/db_xref="GI:7189670"
/translation="MGLFHLTLFGLLCSPILVAKFPESVGHKILYISTGQOAL
AYLEALDAVGDHDFVLRKIGEDYILKQSIHSSDPQTRKSTIIAGIAGSSEALDVS
QAMETADPLDQLLVLSAVSGLGKTSDDLLFKLASYPVIRLEAYRLANLNKTVI
DLHLSFIRLKPBEIIOCLSAIFLRTSESDAYIRDLAKKASIAISAVTLOIGEQO
KRPLPLRLNLTASPODPAIIYALGKIKGDSGYNNIKKOLQPPDVYTLAAAOIL
ALGKEDCDPLVKRKQALREPRALVLRHPSIGIGPIALPIKTKNSKAPKANVLA
LLELCDDTPKLELEYTBRKQPHYNETILASFSKGRLOMKKVVNIIVDPDPERRL
LSTTRGLEOITLFLRLPREAYLPCTIKLASQKTOALATLAFUSHSTHOBALDL
FOAKLPGBEIIIRAVADLAIVNLTQDPEKRSADYAKKLIQETLLEFVDENORPHS
MPYLRQVTPESRTRKIMLDILETLATSKSSSDIRLLIQLMTEGDAKNFPVLAGLLIKI
VE"
6003. .8117
/gene="CP0756"
6003. .8117
/gene="CP0756"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AF38559.1"
/db_xref="GI:7189671"

HDDPCTHNSVSGCYGII.SDSMSYETADIVENPMMLKKNPGFVTCNDWNTYFEGVLT
LSKVAKSPQMANOQLEPYLTLLRQPIRSLITYGTGVIENLVAIVEGQKAPRASHDNTYVENSS
LRYKAPKAKHKVNPRLPGLTSLTSGSLIYTSOVPRISSKHQSLSAKLDQDYFELFK
SVTORHNTLPEPTVLTETPTSLSDKSDHYIIFS.IQDAFHSINTLAKGIDYVLSKTNPR
FRIKAKHTLTTLTSLWESKRPFPKACSESLPFCKKSNVSLDAEWMKIKHSDNMRI
RHEWJGNDVAMTLESLSHRSKSLIKCDRENFILDSRPIDOLLSPLSHRLIIGK
LEVRPHPCNVAITSLRKYGNHODTPMYLEQMLIKGRIPEHMDLYGYERKREDSNPF
FFLKLDKRPKPPF"

complement(11484, .13190)
/ gene="Cp00021"
complement(11484, .13190)
/ gene="Cp00021"
/ codon_start=1
/transl_table=11
/product="Predicted OMP [leader peptide]"
/protein_id="AAD18174.1"
/db_xref="GI:4376273"
/translacion="MGLFPLTLEGLLQSLPISLAKKFPESVGHKILITYTOSTQOAL
ATTYLRALDQVGHDPFVRLKLTGEDYIKQSIHSSDPQTRKSTIIGAGLAKSESLDYTS
QAMETVRLDQQLLVLSAVSGHLSKTSDDLLEFKLASPYPIVRLAARLNLKNTKYV
DHLHSHLHMLPEISQDSALFLRLTESDAYIRQLAAKKSIRSAVLQIGEYQOQ
KKEFLPRLTSLASQDDEALFLRLTESDQSGSYNIKKOKLPDVTYVLAQAOLIL
ALGEDDAPLVKLOEERPRALVLRHNPISIGIPALPIFLKTSQAKLVNLA
LLEIGDMPKLEIYTERIYVQNYNTLSPFGKTIQVQKRWVNIIPDQORETL
LSTRLEEQITFLFRRLKREALPCTIKLASQKQYOLATTAISFLSHTHQEMDLIL
MOLAKIPGPRIIRAVADALVNTLQDPEKRSR.LHDYAKKLTQELFVLDVENDQPHPS
PFLRYQVPESTKTLMLDLLEFLATSKSESDLRLLQMTEDAKNFPVLAGLLIKRI
VE"

complement(13172, .13762)
/ gene="maf"
/ note="Cp00022"
complement(13172, .13762)
/ gene="maf"
/ codon_start=1
/transl_table=11
/product="Maf-type Protein"
/protein_id="AAD18175.1"
/db_xref="GI:4376274"
/translacion="MSLPIPLVIGSSSPRRKFLERFRVFTVIPSNPDESKVSYSGDEPI
ATYQELASAKAAVAVSLHSPDCDILTGDTIVSYDGRIFTRPODKADAIOMLKLRLMDQ
THDYVLSIAVHKGRLTGTSETSOILTFMTPDHRIESYIDTGVTLNCGAYDVCHGGIL
ILKKNHGVYVNOGDELQTLKYLLEELNDLMQST"
complement(13791, .15383)
/ gene="yjjK"
/ note="Cp00023"
complement(13791, .15383)
/ gene="yjjK"
/ codon_start=1
/transl_table=11
/product="ABC Transporter Protein Appase"
/protein_id="AAD18176.1"
/db_xref="GI:4376275"
/translacion="MSVILADTICKSLGSTRLLPDDVSVVENPQNGVGLGPGACKSNTLL
LKIIMKITEPTKRSISLPRKVGGLLRONIDSFHTTYLADCYIMKTRLMELALORDNIXL
LDEFTVATMGELIEELIGENGZYRADSAEELTGIGIPNEKFDKMAAIPIDLOF
RYLFCALGHEPALLLDEPTNLDIYSSIMLNFLKDYEGTVVSHVDFKMTLITTT
HIADIDYDITLIPYGVNDVKEKTKASREOEKADIKSEKKISOLKFAFGAGSVA
SOVSURLREKIKQDOKKSNIOPIRPLSPDSSGKVLSEATIKDYGDAVH
PSELTLYQGDKIGIENNIGKTKTINKLAGVAPSSGSIKIGHOALICSPRQHSQV
LADQGETLEPMNRKRTGIDNDEISVLRKALFPGGDGAKQIALSGSEGTARLLMAQV
MLLENNVNLEIDANNHLDESVSALSALMDYKGTAFVSHQRIQDQCATKLLIFED
KKIITFFDGMVDYTAGHKQL"

complement(15438, .16376)
/ gene="xerc"
/ note="Cp00024"
complement(15438, .16376)
/ gene="xerc"
/ codon_start=1
/transl_table=11
/product="Integrase/recombinase"
/protein_id="AAD18177.1"
/db_xref="GI:4376276"
/translacion="MASTISYFLDYIKMVSASAPHTLRNYCLDNLGILKPIEERGNLA

PSPLOLATEKRXVSELPESLFTKEHVAMTAKLIENKARTIKRCLSSIKSPAHVC
VIOKILENPEETIHGPRLEKELSPMTYAOEVLMAPIDLSKHGLDCLMELFYS
SGNISEIVAVNKODFDLSTHLIRGKKEKRIIPVSNMLOWOILNHPDRKRE
KDPQALFNREGRIRSTRIDRSFOEYLRBSGLSHLTPHRIHTIATHEMESGMDLK
TIOALFLNHSSETTYOVSVKLKOTHRQAHBPHA"
BASE COUNT 4710 a 3330 c 3665 g 4743 t
ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 16448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtggttgatttgaagaaagtcctgagatgtgttataatgattcaaggctccctacc 60
Db 13290 GTGGCTTATTTGAAAAAGGCCATGATGTTTATTAATGTAAGGTCTCCCTATCC 13231
Qy 61 aaactgaataactgtctgagagagttgaacatcgatctatggagctatccatctaac 120
Db 13230 AAACATTAATTAATCTTGTAGAGAGCTTGAACATCGATCTATGGACTATTCATCTAAC 13171
Qy 121 tctcttggaacttattatgtgtgtagctctcccaattctctgtgtgtaaatccctgagtc 180
Db 13170 TCTCTTTGGACTTATTAATGTAGTCTTCCCATTTCTCTTGTGTGTAATTCCTGAGTCC 13111
Qy 181 tgtaggtcataagatcccttataataagtaacgaatctcaacagcagccttagaacata 240
Db 13110 TGTAGGTATTAAGATCCCTTTTATTAAGTAACGAATCTACACAGCGGCTTGAACAATA 13051
Qy 241 tctggaagctctagatgtctgacagtgatcatgaactctctgtgtttaaagaaatctggaga 300
Db 13050 TCTGGAACCTCTAGATGCTCTAGAGGATCATGACTTCTTCGTTTAAAGAAATATGGAGA 12991
Qy 301 agactatctcaagcaaaagcatccactctcagatccgcaaaactagaagaaagcaccatcat 360
Db 12990 AGACTATCTCAAGCAAAACATCCACTCCTCAGATCCGCAAACTAGAAAAAGCACCATCAT 12931
Qy 361 tggaggaagccttgaggagattcttgaagagccttggagctgtctcccaagctatgaaac 420
Db 12930 TGGAGCAAGCCTGGCGGATCTTCAGAAAGCTTGGACGTGCTCCCAAGCATATGGAAC 12871
Qy 421 tgcgaacccctgcagcagctactgtgttttalcgagcagctcgaagacatcttggagaaac 480
Db 12870 TGCAGACCCCTGCAGCAGCTACTGTTTATCGGCACTCTCAGACACTCTTGGAAAAAC 12811
Qy 481 tctggaagctactgttttaaagctttagacatccctcactccgtcaltccgcttagaagc 540
Db 12810 TTTCTGACACTTACTGTTTAAAGCTTTAGCATCTCCATCTCCTGTCATCCCTTGAAGC 12751
Qy 541 cgccatagacttgcttaatttgaagaacactaaagtcattgatactacattcttcat 600
Db 12750 CGCCTATAGACTTGCTAATTTGGAAGAACATTAAGTCAATGATCAATCTTTTAT 12691
Qy 601 tcatatagcttccgaaagaatccaatgcttacttgcggaacatattccagcttggagac 660
Db 12690 TCATTAAGCTCCCGAAGAAATCCAATGCTATCTGCGGCAATATTCTCACCTTGGAAAC 12631
Qy 661 tgaagaatctgagtgcttataatctggagatctctagctgccaagaagcgcgattcgag 720
Db 12630 TGAAGAATCTGATCTTATTTTCGSGATCTCTTAGCTGCCAAGAAACGCGATTCGAG 12571
Qy 721 tgcacagacttgagatcggaagaatccaaacaaagccttcttccgaacaattagaa 780
Db 12570 TGCCACAGCTTTGAGATCGGAGATACCAACAAAACGCTTCTTCGCAACTTGAAGAA 12511
Qy 781 ttgtctaacgagtgcttctcccaagatcaagaagctattctttagcttcaaggaaact 840
Db 12510 TTTGCTTAACGAGTGTGCTCTCTCAAGATCAAGAACCTATTCTTATGTTTAGGGAAGCT 12451
Qy 841 taaggaatgctcagagctactacaatataaaaaagcaattgcagaagcctgagtgtgact 900
Db 12450 TAAAGATGCTCAGAGCTACTACATATATAAAAGCAATTGCAGAAAGCTGATGTGATGT 12391

Qy 901 caacttagcagcagctcgaagctttaaattgtcttggggaagaagagcagctctccgct 960
Db 12390 CACTTAGCAGCAGCTCAAGCTTTTATTTCTTTGGGGAAGAAAGAGAGCTCTTCCCT 12331
Qy 961 gataaaaaagcagcagcttgaggaagcgcctcgaagccctgtatgcttaccgcatcacc 1020
Db 12330 GATAAAAAGCAAGCACTTGAAGAGCGGCTCGAGCCCTGTATGCTTACGCGATCTTACC 12271
Qy 1021 ctctgagatagggatctccgacttgccttcggaatattccctaaacaaactaaagaacggaagc 1080
Db 12270 CTCTGAGATAGGAGATTCCGATTTGCCCTGCCGATTTCCATAAAAGCTAAGACAGCAAC 12211
Qy 1081 caagttgaatgtagctttagctctctttagagtggtgtgacacccctaaactactgga 1140
Db 12210 CAAGTTGAATGTAGCTTTAGCTCTCTTATAGATTAGGAGGTGTGACACCCCTTAACCTACTGGA 12151
Qy 1141 atacattacgaagaagctgtgtccaaaccaataataatgagactcagacttgaagttcttc 1200
Db 12150 ATACATTACCGAAGGCTTGTCCAAACCATTTATTAATGAGACTCTAGCCTTGAGTTTCTC 12091
Qy 1201 taaggagcgtacttatacaaaatttggaaagcgggtgaacatcatagttccctcaagaatccca 1260
Db 12090 TAAGGGCGCTACTTTACAAAATTGGAGGGGTTGAACATCATAGTCCCTCAAGATCCCA 12031
Qy 1261 ggaagaggaaaggtgtgtctcccaaacccgaggtcttgaagagcagatccttagttctc 1320
Db 12030 GGAGAGGGGAAGGTTGCTCTCCACAACCCGAGGCTTTGAAGAGCAGATCTTACGTTTCT 11971
Qy 1321 ctccgcttaccctaaagaagctctacccctcgtattttaaagctcttggggaagtcaga 1380
Db 11970 CTTCGGCTACTTAAGAAAGCTTACCTCCCTGTAATTTATTAAGCTTTTGGCAGTCAGAA 11911
Qy 1381 aactcaagcttgcacactctgagcttcttctttaaagtcacacactacatcaagaagcctt 1440
Db 11910 AACTCAGCTTGCCACTACTGCGATTTCTTTTAAATGTCACACCTCAACATCAGAAAGCCTT 11851
Qy 1441 agatctacttcttccaagcttggaagcctctcccttggaaacatcatcccgagctatgaga 1500
Db 11850 AGATCTACTTTTTCCAAGCTGCGAAGCTTCCGGAACCTTCAATCAATCCGCGCTATGACA 11791
Qy 1501 tcttgctatttaacttcaacccaagaatcccttgaagaaagcttcttccatgattatgac 1560
Db 11790 TCTTGCTATTATTAATCTCACCAGAAATCTGAAAAAAGCTTCTTCATGATTAATGAC 11731
Qy 1561 aaaaaagctaatctcagaaaccttgttatttltgtgacacggaagaaacccaagacccatcc 1620
Db 11730 AAAAAAGCTAATTCAGGAACCTTGTATTGTGAGACAGGAACCAAAAGACCCCATCC 11671
Qy 1621 cagcagtcctatactagcttataagtgtaacccccaagaagccgttagaagctcatgttga 1680
Db 11670 CAGCATGCTCTATCTTACGTTTATCAAGGTCAACCCCAAGACCGTACGAAGCTCATGTTGGA 11611
Qy 1681 tatctagaagaacactagcacaactccgaagcttcttccgaagatatacgttlatatgataact 1740
Db 11610 TATTTTAAAGACACTAGGCACTCGAAGTCTTCGAAAGATATCCGTTTATTTATATCAACT 11551
Qy 1741 gatgacggaaggaatgtcaaaaaatltccagctctctgcaggtctactcatataaaaaattgt 1800
Db 11550 GATGACGGAAGAGATGCAAAAAAATTTCCAGTCTCTCAGAGCTTACTCATTAATAATTGT 11491
Qy 1801 ggagtaaccccaactagcttataagaagcttcttcttatttccatctcccttctgttc 1860
Db 11490 GGAGTAACCCCAACCTTACGTTTATGAAGCGTGTCTTATTTCTTACCTTCCCTTGTTC 11431
Qy 1861 ttatggttctcagctgagcttggacttcaatcaagaagcgtgtgaa 1907
Db 11430 TTAATGCTTCTCAGCTGATGCTTTGACTCATGAAGAGCTGTGAAA 11384

RESULT 4
AX349527
LOCUS AX349527 1707 bp DNA Linear PAT 06-FEB-2002
DEFINITION Sequence 50 from Patent WO0202606.

ACCESSION AX349527.1 GI:18615370
VERSION
KEYWORDS Chlamydia pneumoniae.
SOURCE Chlamydia pneumoniae.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
1 (sites)
AUTHORS Ratti, G. and Grandi, G.
TITLE Immunisation against Chlamydia pneumoniae
JOURNAL Patent: WO 0202606-A 50 10-JAN-2002;
Chiron S.P.A. (IT)
FEATURES
source Location/Qualifiers
1.1707
/organism="Chlamydia pneumoniae"
/db_xref="taxon:83558"
BASE COUNT 493 a 415 c 340 g 459 t
ORIGIN
Query Match 89.5%; Score 1707; DB 6; Length 1707;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 101 atgggaacttccatctaaactctcttggaacttatttgtagtcttcccatcttctt 160
DB 1 ATGGGACTATTCATCTACTCTCTTGGACCTTTATGTGTAGTCTCCATTTCTCTT 60
QY 161 gtgtgcaaatctccgtgaactgttagtcaataatccttataatgaatgaactaa 220
DB 61 GTTGCTAAATTCCTGAGTCTGTAGTCAATAGATCCTTATATATAGTACGCATCTACA 120
QY 221 cagcagccttagcaacacatctggaagctctagatcgctacggtgtagatgaactctc 280
DB 121 CAGCAGCCTTAGCAACATATCTGAGAGCTTAGATGCTACGCTATCATGACTCTTC 180
QY 281 gtcttaagaanaaatcggaagaactatctcaagcaaaagcatccactcctcaatccgcaa 340
DB 181 GTTTTAAGAAAAATCGAAGACATCTCAAGCAAGCATCCACTCTCAGATCCGCAA 240
QY 341 actaagaanaagcaactgtaggaagcgtggaagcgtggaagcgttcaagaagcgtggaagcgtg 400
DB 241 ACTAGAAAAAGCACCATCTGTGAGGAGGCGCTGGCGGANTCTTCAAGCCTTGGAGCTG 300
QY 401 ctctcccaagctatggaactgagaccccgctgcaagcagctactggtttatcggagctc 460
DB 301 CTCTCCCAAGCTATGAAACTGAGACCCCTGACAGCTACTGGTTTATGCGAGTGC 360
QY 461 tcaagacatcttgggaanaactctgaagactactgttttaagaactttagcatctccat 520
DB 361 TCAGGACATCTTGGGAAACTTCTGACGACTTACTGTTTAAAGCTTTAGCATCTCCCTAT 420
QY 521 cctgtatccgcttagaagcccgctataagctgtcaatttgaagaacataaagctatc 580
DB 421 CCGTGATCCGCTTAGAACCCGCTTATAGACTTGTGAATTTGAAGAACATTAAGTCATT 480
QY 581 gatcatcatatctcttcaacttaagctcccggaanaatccaagcctatctgggca 640
DB 481 GATCATCTATCTTCTTCAATCTAAGCTTCCGGAAGAAATCCAATGCCATATGCGGCA 540
QY 641 atattctcaagcttggagactggaanaatctgactatataatcgggactcttagctgc 700
DB 541 AATATCTCAAGCTTGGAGACTGMAAATCATGATGATTTATTTGGGATCTCTTAGCGCC 600
QY 701 aagaagaagcgagatctggaggtgacagacttgcagatcgggaataatacaaaaaagc 760
DB 601 AAGAAAAAGCGGATTCGGAGTGCACAGCTTTCAGATGGGAGAAATACCAAAAAAGCGC 660
QY 761 ttcttcgcgaacttaggaattgtccaagagtgctctcctcaagaataaagaactat 820
DB 661 TTCTCTCGAGCACTTAGGAATTTGCTACGAGTGGCTCTCTCAAGATCAAGAACTATT 720
QY 821 cttatgctttaggaagacttaagatggtcagagctactacaataataaaaaagcaatg 880
|||||

DB 721 CTTTATGCTTTAGGAAGCTTAAGATGGTCAGAGCTACTACATATATAAAAAACAATTG 780
QY 881 cagaagcctgagtgtgagtcaactttagcagcagcccaagcttaattgcttgggga 940
DB 781 CAGAAGCTGATGTGATGTGCACTTTTAGCAGCGCTCAAGCTTTAATGTGTTGGGAAA 840
QY 941 gaagaggaagccttccgtgataaaaaagcaagcaactttaggaagcgagcctcgaagctg 1000
DB 841 GAAGAGGAGCGCTTCCCGTGATAAAAAGCAAGCACTTGAAGAGGCGGCTCGAGCCCTG 900
QY 1001 tatgctttagcagcatctaccctctgagaagaatccgattctgcctgcgataatccca 1060
DB 901 TATGCTTAGCGCATCTACCTCTGTGAGATAGGATTCGATTCGCTCCGTGATATTCCTA 960
QY 1061 aaacactaagaagcagcagcagcagcttgaatgttagctttagctctctttagagttg 1120
DB 961 AAACCTAAGAACGCGAAGCAAGTTGAATGATGCTTTAGCTCTCTTAGAGTTAGGCTGT 1020
QY 1121 gacacccctaaactactggaataacataccgaagaagctgtgccaaacataatag 1180
DB 1021 GACACCCCTAAACTACTGGAATACATTACGGAAGGCTTGTCCACACATTTAATAG 1080
QY 1181 actctagccttgagttctcttaagggcgctacttcaaaaattggaagcggtgagacatc 1240
DB 1081 ACTCTAGCCTTGAAGTTCTCTTAAGGGCGCTTACAAAATTGGAAAGCGGTGAACATC 1140
QY 1241 atagtcacctcaagatcccgagagaggaaggttgctctccaaacccgaagcttgaa 1300
DB 1141 ATAGTCCCTTAACATCCCGAGAGGAGGAAGTTGCTCTCCACACCGGATCTTGAA 1200
QY 1301 gaagagatcccttagcttctctctccgactcaactaaagaagcttaccctccgttattat 1360
DB 1201 GACACAGTCTCTTACGTTTCTCTCCGCTTACCTAAAGAACCTTACCTCCCTGATTTAT 1260
QY 1361 aagctttagcgagtagaagaactcagctgcacactgagatcttctttaaagtaac 1420
DB 1261 AACCTTTGGCGAGTGAAGAAACTCAGCTTCCACTACCTGATTTCTTTTAAAGTCA 1320
QY 1421 acctcaatcagaagaagccttagatctacttllcaagcttgcgaagcttccgtgagaacct 1480
DB 1321 ACTCATATGAGAGGCTTAGATCTTACTTTTCCAAAGCTCGAAGCTTCTGGAGACCT 1380
QY 1481 atcatccgagctatcagatctctgtattatataatctcaacaaagaatccctgaaaaaaa 1540
DB 1381 ATCATCGCGCCCTATGAGATCTTGCATTTAATATCTACCAAGATCTGAAAAAAA 1440
QY 1541 cgttctccatgattatgcaaaaaagctaaatcagaagaaccttgtatttggagacg 1600
DB 1441 CGTTCTCTCATGATTTATGCAAAAAAGCTAATTCAGGAAACCTTGTATTGTGACACG 1500
QY 1601 gaaacccaagaagcccatcccaagcagcctatctagatgaagtcacccagaagc 1660
DB 1501 GAAACCAAGAGCCCAATCCACAGATGCCCTATCTACGTATAGGTCACCCAGAAAGC 1560
QY 1661 cgtcagaagctcatgttgaatatcttagagacactgacacccctcgaagcttccgaagat 1720
DB 1561 CGTACGAGCTCATGTTGATTTCTTAGAGACACTGACCCACCTCGAAGTCTTCGAAAGAT 1620
QY 1721 atccgcttattgatacaactgtagcaggaagaagatgcaaaaaattccagctcttga 1780
DB 1621 ATCCGTTTATTTGATTAACAAGTACGAGGAGGAGATGCAAAAAATTTCCAGTCTTGCA 1680
QY 1781 ggccttactcaataaaattgtagagtaa 1807
DB 1681 GGCTTACTCATTAATAAATTGTGAGTAA 1707
RESULT 5
AX349569
LOCUS AX349569 2088 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 92 from Patent WO0202606.
ACCESSION AX349569
VERSION AX349569.1 GI:18615391

KEYWORDS	Chlamydomophila pneumoniae.
SOURCE	Chlamydomophila pneumoniae
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE	1 (sites)
AUTHORS	Ratti,G. and Grandi,G.
TITLE	Immunisation against Chlamydia pneumoniae
JOURNAL	Patent: WO 0202606-A 92 10-JAN-2002;
FEATURES	Chiron S.p.A. (IT)
Source	Location/Qualifiers 1. .2088 /organism="Chlamydomophila pneumoniae" /db_xref="taxon:83558"
BASE COUNT	574 a 480 c 417 g 617 t
ORIGIN	
Query Match	4.4%; Score 84; DB 6; Length 2088;
Best Local Similarity	100.0%; Pred. No. 2.3e-37;
Matches	84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1824 atgaacgttgctctctattctctagctcctctgtctctatgaggtctcctcagctatgct 1883
Db	1 ATGAACGTTCCTCTATTCTCTAGCTCCCTTGTCTTATGGGTCTCCTACGCTATGCT 60
QY	1884 ttgactcatcaagagcgtctgtaa 1907
Db	61 TTGACTCATCAAGAGGCTGTGAAA 84
RESULT	6
LOCUS	AX100531/c 36 bp DNA linear PAT 10-APR-2001
DEFINITION	Sequence 4 from Patent WO0121804.
ACCESSION	AX100531
VERSION	AX100531.1 GI:13619535
KEYWORDS	synthetic construct.
SOURCE	synthetic construct.
ORGANISM	artificial sequence.
REFERENCE	1 (bases 1 to 36)
AUTHORS	Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
TITLE	I chlamydia /I antigens and corresponding dna fragments and uses thereof
JOURNAL	Patent: WO 0121804-A 4 29-MAR-2001;
FEATURES	Aventis Pasteur MSD (FR)
source	Location/Qualifiers 1. .36 /organism="synthetic construct" /db_xref="taxon:32630" /note="3' PCR primer"
BASE COUNT	8 a 12 c 7 g 9 t
ORIGIN	
Query Match	1.3%; Score 24; DB 6; Length 36;
Best Local Similarity	100.0%; Pred. No. 0.066;
Matches	24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1781 ggcctactcataaaatgtgag 1804
Db	36 GGCTTACTCATTAATAATGTGAG 13
RESULT	7
LOCUS	AX100530 44 bp DNA linear PAT 10-APR-2001
DEFINITION	Sequence 3 from Patent WO0121804.
ACCESSION	AX100530
VERSION	AX100530.1 GI:13619534
KEYWORDS	synthetic construct.
SOURCE	synthetic construct.
ORGANISM	synthetic construct

REFERENCE	artificial sequence.			
AUTHORS	1 (bases 1 to 44)			
TITLE	Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.			
JOURNAL	I chlamydia /1 antigens and corresponding dna fragments and uses thereof			
FEATURES	Patent: WO 0121804-A 3 29-MAR-2001;			
source	Aventis Pasteur MSD (FR)			
	.location/Qualifiers			
	1..44			
BASE COUNT	12 a 14 c 8 g 10 t			
ORIGIN				
Query Match	1.3%: Score 24; DB 6; Length 44;			
Best Local Similarity	100.0%: Pred. No. 0.066;			
Matches 24; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
OY	101	atggaactatccatctactctc	124	
DB	21	ATGGACATATCCATCTACTCTC	44	
RESULT	8			
AE001308				
LOCUS	AE001308	11627 bp	DNA	linear BCT 30-OCT-2000
DEFINITION	Chlamydia trachomatis section 35 of 87 of the complete genome.			
ACCESSION	AE001308	AE001273		
VERSION	AE001308.1	GI:3328766		
KEYWORDS				
SOURCE	Chlamydia trachomatis.			
ORGANISM	Chlamydia trachomatis			
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
AUTHORS	1 (bases 1 to 11627)			
	Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,			
	Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,			
	Koonin,E.V. and Davis,R.W.			
TITLE	Genome sequence of an obligate intracellular pathogen of humans:			
JOURNAL	Chlamydia trachomatis			
MEDLINE	Science 282 (5389), 754-759 (1998)			
PUBMED	99000809			
PUBMED	9784136			
TITLE	2 (bases 1 to 11627)			
JOURNAL	Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,			
MEDLINE	Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.			
PUBMED	Comparative genomes of Chlamydia pneumoniae and C. trachomatis			
PUBMED	Nat. Genet. 21 (4), 385-389 (1999)			
PUBMED	99206606			
PUBMED	10192388			
TITLE	3 (bases 1 to 11627)			
JOURNAL	Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,			
MEDLINE	Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,			
PUBMED	Koonin,E.V. and Davis,R.W.			
PUBMED	Direct Submission			
TITLE	Submitted (20-MAY-1998) Program in Infectious Diseases, University			
JOURNAL	of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA			
FEATURES	.location/Qualifiers			
source	1..11627			
	/organism="Chlamydia trachomatis"			
	/strain="D/UW-3/CX"			
	/db_xref="taxon:813"			
	207..1121			
gene	/gene="eIaC"			
	/note="CT346"			
	207..1121			
	/gene="eIaC"			
	/codon_start=1			
	/transl_table=11			
	/product="Subphosphatase/Glycosulfatase"			
	/protein_id="AAC67941.1"			
	/db_xref="GI:3328767"			
CDS				

gene
1183. .2130
/gene="xerc"
/note="CT347"
1183. .2130
/gene="xerc"
/codon_start=1
/transl_table=11
/product="Integrase/recombinase"
/protein_id="AAC67942.1"
/db_xref="GI:3328768"
/translation="MTSPYAFDLDTLKNKASLHTLRNYCDLSLCKFLKKKSDLS
PTPLSLDNTYDPLPSFLTKDNIRLYLEQIQTHSKRTVRRRSAIKSFARFC
VKNQLIPEPAMRIGPRLPOLPSPLYEVLALMAAPELDKVGFPRDCLLELYS
SGLRISEITALNRADIDFOSHLHIRGKGERIYPMTRVAQMLQDYLNHDDRAVE
ODHACFLNRFGRRLSTRSIDRKFOOYLTKTGLSGSITPHRTIATHWLERGMDLK
TIOLLGHTSLFTTTIYTHVSMKLLKKQIHDEHPHNLK"
2184. .3770
/gene="yjkK"
/note="CT348"
2184. .3770
/gene="yjkK"
/codon_start=1
/transl_table=11
/product="ABC transporter Protein ATPase"
/protein_id="AAC67943.1"
/db_xref="GI:3328769"
/translation="MSIVLDKIGKTLGTPRVLPFDVSVFNPNGNRYGLTNGAGSKTL
LKITGIIVPSRGTISLPKIGILRONIDSGDYVLDGVIMGNSRIMAMGRALY
AEFFDAIGMELGEIEBITGEDGYRAESEAEELLGIGIPEEPFSQKATTPLDIQF
RYLLIQALGHPREALLDDEPTNHLDLHSLNWLGSFLKDYNGTVIVSHDRHFLNITTT
HLADIDYDTIITIPGNDAMVEMKTAHSROEKADIKSEKKIAQLREFYKAGSASRA
SOVQRLREIKIGPOELKSNIOBPYIRFVPSDKASGVPSIEGISITPADPLFK
PSLEIYOGDKIGVIGNNGIKGTTLMKILLAGVEYDSOGQVKGHNVAESYFPONADI
LKDCGEFTLEWLRRRKGTIPDOEYRSVIGKMLFGDDDAFKOVKALSGETRRLMAG
KMLENNHMLIDEANNHLDLESVSALAMAINDPGATLIVSHDRTLIDECATRLIFN
KGTISFGGPMSEYLAOSK"
3806. .4396
/gene="maf"
/note="CT349"
3806. .4396
/gene="maf"
/codon_start=1
/transl_table=11
/product="Maf-type protein"
/protein_id="AAC67944.1"
/db_xref="GI:3328770"
/translation="MEARLYLGSSSESRKAVLESRIPIFCVSPDFDENSTIYSGDP
EYTKELAMKANAVNSQGSFALITADVVVYKGVFNKPESEEHAVMLPTLSGSS
HSVITTLVLMONEKVLASSENTQVSFIDIPQHLKTYVRSFSLKRCGGYCVQDGGGL
IKQIEGCVYNIQGLPIKTLNQLMELFNISMDYLV"
4378. .6078
/gene="CT350"
4378. .6078
/gene="CT350"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67945.1"
/db_xref="GI:3328771"
/translation="WGLSRLATISFLSFTLSASCDPSPSSVSQRILFSCRSKSVQAIEA
YLEASATYQOHDFSVLYIAESYLOQSFSEDTYIRKSAIIGAGLSGSEALELSEA
IEFDOLYEDLIINATSQLSKTSKSLFGLTASHPIYRLAAGYLAGMKASVSDY
LVSFYIKLPEEIONLAATIFLOLEADAYIHLHLSPPNLTNRNVAVLIGEYKQK
FLPTLRSLITSASPLDEGALAIKGLDNGSYPRIKALSRSNPENVAAAOITLFL
EKKEEALPITNLCOOKILRALYTAFLSQEGBELLPIFYNATDELRINTALAV
HOGCTDPOVLAHTLEILSKVLHRIPLPHSTGKAIQFWKECTTPPLMSQDEKMTLA
MYVAEDTILSALKIPNDAYLPIERILIASOKTIIAANAIAFLSVTAHPQALISVSK

gene
6085. .8178
/gene="CT351"
6085. .8178
/gene="CT351"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67946.1"
/db_xref="GI:3328772"
/translation="MKREFPIGVLAHTLPSEGLSHQOAVQKISYLSHFQITGI
MVEDGVLIHDDLRLOANKAVENRTDGIKIVAGNVAVRGKILICDYLEVED
TDSLCTNGRCSLYPMFTIGSTITISPSIITHKGIISTSEGPKHILCSDYLYKSS
DSVLMSGPSRLSICMPVULLPOISIMEPEIKPPTIFRSGSGGFLGYSYSPIS
KHCSTTLPDGFRRHGIGLGVNMRFSSEENSNAINIKSYIAHRLAIDSSAKDRY
LHGDFTFSEKRAHLAGEPFLSDSWETVVDIFPNNSLAKVTGEVLSWRDNNLFCKM
TSSVYNSFQNVKQELPOAILHHRVPIARRSRIPIENRLEAGFLDFHFSNINSGNFS
SWRFSNAHVYGLVLIPTGLTSPSLSGTAIYTYRMLSPNAACOLSGSLSPDYRVALQ
KEYRAHRHIVEPESFLKTRRVLSDEPHIISIDAFHSINLHVGLSEKYLANKST
PSHLKMTPTYIDDEPHAKDTPFKTACMSPLEPTLQNTISLDAEMTKSRMHNVIY
EWILNDNLGLIETLHRSKYGTICAKDNYTLDVSSGLDTLILASPLSDRRNLTGKLF
VRRPHMNYNMLRYGWRHPDPSYLEYQMLIGHKIFEMQLFSYKREKADKRCFEY
LKLDRKQKRRHPCG"
8341. .8646
/gene="CT352"
8341. .8646
/gene="CT352"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67947.1"
/db_xref="GI:3328773"
/translation="MRLNLTAKKILQLTRLPFSKFLGILIOINRNRHKKTLTIPC
SYSLQEQVHTLSITKGIIRTKCKTEQPPYPPKILGERTKLALNPTFCHSVA"
complement(8716. .9261)
/gene="def"
/note="CT353"
complement(8716. .9261)
/gene="def"
/codon_start=1
/transl_table=11
/product="Polypeptide Deformylase"
/protein_id="AAC67948.1"
/db_xref="GI:3328774"
/translation="MIRDEITYDSPLRKVAVPEITDELQVLDNSETMAFYKGV
GLAAPQVGQISILFTMGVERLEDELVPCDFPRVFINVIQKSEQALVYNGECIST
PGLRGEVARPDKITVSAKNLDQQLSALGFLARIYMHETDHLGVLYIDRMSDK
TKQFNMLEKIRRKYSILRGL"
complement(9536. .10369)
/gene="kgsA"
/note="CT354"
complement(9536. .10369)
/gene="kgsA"
/codon_start=1
/transl_table=11
/product="Dimethyladenosine Transferase"
/protein_id="AAC67949.1"
/db_xref="GI:3328775"
/translation="MARSSIEQLTSFLRSVNGRAKKALSONFLVDGNILTKILTTAAV
OPGDWVLEIGPGFALSEVLLSQGANVIALEDPMPEESLSQLEPMOIEITDACKYPLT
SLDEKGMKGRIVANLPHYITTPLTCKFLPCPYPMKVIYTWIODEVARITAKGD
KYGSLTFTVLSFPAFOVAFKYSPNCFYKPSVSHSVAMRVHEDQALADSELEFPET
LTPRAFGORRKLANSKMLYPRKDYFOVLQDLGSEKTRPETIFLEETLIKPHLLKD
I"
complement(10385. .11446)
/gene="CT355"
complement(10385. .11446)
/gene="CT355"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

/protein_id="AAC67950.1"
/db_xref="GI:3328776"
/translation="MRVERAVIACVITITFSGIARCEGAFSSGSEPNQSGVAIH
NRVFKVDEDTVVTTLDVTRKNTLITFSTCPQVDSVASRQYVSAMPVLETVINE
FLMADAKAKKIFIDIPNSVNOEIAMFGKRD,SPPAKFEPMDEPVNVNHLIAQRY
EGMVRSRVMLKVTGMYREYQKLAEEAQITQMTYVLTIKAGLEFLARKIKQVO
ERLNEGSSMDKERLTALVLSOGOLMCSEEFREDADQSVNHRQSLERINPERCKG
VLEHAGSLKLEVLNRAKTLDPDKMEAOIKQOLMEFAAEENANKNKIHAHYGFD

Query Match      1.2%: Score 22; DB 1; Length 11627;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 519 atccgtcatccgcttagaagc 540
|||||
Db 4790 ATCCTGTCATCGCTTAGAGC 4811

RESULT 9
AC017438 47445 bp DNA linear HTG 09-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
DEFINITION pieces.
ACCESSION AC017438.1 GI:553548
VERSION AC017438.1
KEYWORDS HTG; HTGS; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 47445)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDL:10210763 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..47445
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 14508 a 9495 c 8973 g 14469 t
ORIGIN
Query Match      1.2%: Score 22; DB 2; Length 47445;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 860 tacatataaaagcaattgc 881
|||||
Db 34100 TACAATATATAAAGCAATTGC 34121

RESULT 10
AC009341 183722 bp DNA linear INV 10-MAR-2001
LOCUS Drosophila melanogaster, chromosome 2L, region 36C-36C, BAC clone
DEFINITION BACR07G13, complete sequence.
ACCESSION AC009341
VERSION AC009341.6 GI:13270516
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 183722)
REFERENCE
1 (bases 1 to 183722)

```

```

AUTHORS
Celisner, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Bazon, J., Beson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Chame, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnak, D., Farfan, D.,
Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwan, C., Jalali, M., Kruse, D., Li, P., Matel, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Pacled, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Swirskas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 2L, region 36C-36C
Unpublished
2 (bases 1 to 183722)
REFERENCE
Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Chame, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Katra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomolan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Swirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
JOURNAL Submitted (17-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 10, 2001 this sequence version replaced gi:7264746.
COMMENT Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
FEATURES
source
1..183722
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="36C-36C"
/clone_id="BACR07G13 (D1029)"
/clone_id="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
PBACE3.6)"
BASE COUNT 54821 a 35450 c 37135 g 56316 t
ORIGIN
Query Match      1.2%: Score 22; DB 3; Length 183722;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 860 tacatataaaagcaattgc 881
|||||
Db 50434 TACAATATATAAAGCAATTGC 50413

RESULT 11
AE003654 261000 bp DNA linear INV 04-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386055 section 47
DEFINITION of 63, complete sequence.
ACCESSION AE003654
VERSION AE003654.1 GI:7298369
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

```

REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 261000)
REFERENCE	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanalides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champagne,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abail,J.F., Agbayani,A., An,H.U., Andrews-Plannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Bereman,B.P., Bhandari,D., Bolshakov,S., Borkov,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S., Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,S.D., Nelson,K.A., Nixon,K., Nusser,D.R., Pacle,J.M., Palazolo,M., Plittman,G.S., Pan,S., Pollard,J., Puri,Y., Reese,M.G., Reiner,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,K.A., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissendach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
JOURNAL	The genome sequence of <i>Drosophila melanogaster</i>
MEDLINE	Science 287 (5461), 2185-2195 (2000)
AUTHORS	2 (bases 1 to 261000)
TITLE	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.
JOURNAL	Direct Submission
TITLE	Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
FEATURES	Location/Qualifiers
source	1. 261000
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
mRNA	/chromosome="2L"
	<34860..>35096
	/gene="CG12621"
	/product="CT35033"
gene	/db_xref="FLYBASE:FBan0012621"
	/db_xref="FLYBASE:FBgn0032627"
	/evidence=not_experimental
	<34860..>35096
	/gene="CG12621"
	/map="36B1-36B3"
	/db_xref="FLYBASE:FBan0012621"
	/db_xref="FLYBASE:FBgn0032627"
	/evidence=not_experimental
	34860..35096
CDS	/gene="CG12621"
	/note="CG12621 gene product"
	/codon_start=1
	/db_xref="FLYBASE:FBan0012621"
	/db_xref="FLYBASE:FBgn0032627"

gene	/evidence=not_experimental
	/protein_id="AAF53596.1"
	/db_xref="GI:7298370"
mRNA	/translation="MPEVKIPNHIMKLSATLGCRAALDGSLSYKWKDKDHEIYRYVPRDKPGGSFPLPGVNIIDVSPPODIENKFLIC"
	join(<57505..55822,55985..56151,56624..56726,56970..57585,57644..>57725)
	/gene="CG15137"
	/product="CT35034"
gene	/db_xref="FLYBASE:FBan0015137"
	/db_xref="FLYBASE:FBgn0032628"
	/evidence=not_experimental
	<57505..>57725
	/gene="CG15137"
	/map="36B3-36B4"
gene	/db_xref="FLYBASE:FBan0015137"
	/db_xref="FLYBASE:FBgn0032628"
	/evidence=not_experimental
	join(<57505..55822,55985..56151,56624..56726,56970..57585,57644..>57725)
	/gene="CG15137"
	/note="CG15137 gene product"
CDS	/codon_start=1
	/db_xref="FLYBASE:FBan0015137"
	/db_xref="FLYBASE:FBgn0032628"
	/evidence=not_experimental
	/protein_id="AAF53597.1"
	/db_xref="GI:7298371"
	/translation="MNAKLRLVIEESRLSLQLRNGAFMRCVYLSDOLEDAKTSIYICILITFLYDALVNCNMRLRVLDHKKDFLGLIERVTFASLDIRLESEFYKIMDLSEFLGAVYYGFIETLSNEEDFMNNGRLEFTLLRWYQCKMMDLASKYRMVVARPQVRMSRSGILITKEIFGSIITDGLQMAVLYSAMGSVDSPFLGLGLQWMMYILMMAMQOHIMILFVTRTOLITLRLROYLDEAKDLILSRQGVMTKCCSLADQENIARIOSQLOITINQMEVRFQIGIAMTYGGYIYSSVGTCTLAYSLIKHGENLSKMLSYIILAYSWMITGKCYRSGSGQYSLAMWSASKRP"
mRNA	complement(join(<66248..66528,66595..66730,67132..67397,67923..>68151))
	/gene="CG15138"
	/product="CT35035"
gene	/db_xref="FLYBASE:FBan0015138"
	/db_xref="FLYBASE:FBgn0032629"
	/evidence=not_experimental
	complement(<66248..>68151)
	/gene="CG15138"
	/map="36B3-36B4"
gene	/db_xref="FLYBASE:FBan0015138"
	/db_xref="FLYBASE:FBgn0032629"
	/evidence=not_experimental
	complement(join(<66248..66528,66595..66730,67132..67397,67923..>68151))
	/gene="CG15138"
	/note="CG15138 gene product"
CDS	/codon_start=1
	/db_xref="FLYBASE:FBan0015138"
	/db_xref="FLYBASE:FBgn0032629"
	/evidence=not_experimental
	/protein_id="AAF53598.1"
	/db_xref="GI:7298372"
mRNA	/translation="MPRLTTHNSPPTIGSHHTTARNSPDPADFAGPLANNKSKPPFPFCRRARSIEILNDLRAILITARSDFDLFWVDLPDEGSFKISGGRPRQIGIYVNVNCFAGRSKPAVKLSMVNGVEEQQLRKQDTVSGRDGLSTVLGIQFVQKHFPRKGMKRLCIAELSTVYWRNCNEESVGGDRQAPVSESETVYASNSRADPVOASTASAPSPITPLVLIPVALAVWMAVTLAAGITRDIDDKISMDRPAQGIQITMETQOARELLAGKKEENRRTSHKSAQTOLKIALTPAR"
	<11562..>115853
	/gene="CG15139"
	/product="CT35036"
gene	/db_xref="FLYBASE:FBan0015139"
	/db_xref="FLYBASE:FBgn0040987"
	/evidence=not_experimental
	<11562..>115853
	/gene="CG15139"
	/map="36B5-36B5"

```

/db_xref="FLYBASE:FBgn0015139"
/db_xref="FLYBASE:FBgn0040987"
/evidence-not_experimental
115662..115853
/gene="CG15139"
/note="CG15139 gene product"
/codon_start=1
/db_xref="FLYBASE:FBgn0015139"
/db_xref="FLYBASE:FBgn0040987"
/evidence-not_experimental
/protein_id="AAF35399.1"
/db_xref="GI:7298373"
/translational="MAVINPQWRPASDLRSQSTPRKDVNDAAEEQQEQEHIIISGDS
VSHEDTSIVSGTNSPN"
complement(join(<127732..127918,136653..136803,
136950..137197,138481..138519))
/gene="CG6365"
/product="CT19870"
/db_xref="FLYBASE:FBgn006365"
/db_xref="FLYBASE:FBgn0032630"
complement(<127732..>138519)
/gene="CG6365"
/map="36B5-36B6"
/db_xref="FLYBASE:FBgn006365"
/db_xref="FLYBASE:FBgn0032630"
complement(join(127808..127918,136653..136803,
136950..137197,138481..138519))
/gene="CG6365"
/note="CG6365 gene product"
/codon_start=1
/db_xref="FLYBASE:FBgn006365"
/db_xref="FLYBASE:FBgn0032630"
/protein_id="AAF33600.1"
/db_xref="GI:7298374"
/translational="MAJASVLAGAVANKTVKPYKTEVTHRIKRALRLVLSAGIVATS
VPHFERLDTKNGKMYLKDTEFLPREHNGSOVDWINGVALSVICVCPOTY
DRKOLINSKGTAVKONQVIEIKVFNILGILKKRKNQSLSKSYKISGAGSHKMSA
NNOHWPSPQAFSGSVPHRTER"
join(<162848..162875,162979..>163769)
/gene="CG15140"
/product="CT35037"
/db_xref="FLYBASE:FBgn0015140"
/db_xref="FLYBASE:FBgn0032631"
/evidence-not_experimental
<162848..>163765
/gene="CG15140"

gene
Query Match 1.2%; Score 22; DB 3; Length 261000;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 860 tacataataaaagaattgc 861
|||||
Db 41537 TACAATATATAAAGCAATTGC 41516

RESULT 12
LOCUS AX035433 4343 bp DNA linear PAT 15-NOV-2000
DEFINITION Sequence 8 from Patent EP1035209.
ACCESSION AX035433
VERSION AX035433.1 GI:11191075
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 4343)
AUTHORS Azzey,M., Hobom,G., Wenke,A. and Flick,R.
TITLE Stable recombinant influenza viruses free of helper viruses
JOURNAL Patent: EP 1035209-A 8 15-SEP-2000;
ARTEMIS PHARMACEUTICALS GMBH (DE)
FEATURES
Location/Qualifiers
1..4343

```

```

/organism="synthetic construct"
/db_xref="taxon:32630"
/note="PHI2989"

BASE COUNT 1085 a 1064 c 1100 g 1094 t
ORIGIN

Query Match 1.1%; Score 21; DB 6; Length 4343;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1546 tctcatgattatgcaaaaa 1566
|||||
Db 1876 TCTCATGATTATGCAAAAA 1896

RESULT 13
LOCUS AL670958 177014 bp DNA linear HTG 24-JAN-2002
DEFINITION Mus musculus chromosome 4 clone RP23-410K19, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL670958
VERSION AL670958.1 GI:18376488
KEYWORDS HTG: HTGS_PHSBL; HTGS_ACTIVEFIN.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 177014)
AUTHORS Clarke,D., Connor,R., Leaves,N.I., Caveberry,L., Greystrom,J.,
North,P.C., Hunter,G., Shuflebottom,L., Kimberly,C., Campbell,M.,
Jones,S., Lawrence,N., Strachan,G.L., Greenham,L., Maggott,K. and
Botcherby,M.R.M.
Direct Submission
Submitted (23-JAN-2002) Mouse Sequencing Group, HGMP-RC, Hinxton,
Cambridge, CB10 1SB, UK. E-mail enquiries:- mrbotcherb@hgmp.mrc.ac.uk
or clare2@hgmp.mrc.ac.uk
REMARK HGMP-RC part of the UK Mouse Sequencing Consortium
COMMENT * NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved

FEATURES
Location/Qualifiers
1..177014
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-410K19"
/clone_1tb="RPCT-23"
1..12805
/note="assembly_fragment:1158"
misc_feature 12906..67118
/note="assembly_fragment:201"
misc_feature 67219..99173
/note="assembly_fragment:2057"
99274..111725
/note="assembly_fragment:2060"
111826..115810
/note="assembly_fragment:225"
115911..139424
/note="assembly_fragment:2668"
139525..156165
/note="assembly_fragment:278"
156266..170601
/note="assembly_fragment:4020"
170702..177014
/note="assembly_fragment:4351"
BASE COUNT 52674 a 37604 c 36695 g 49232 t 809 others
ORIGIN

Query Match 1.1%; Score 21; DB 2; Length 177014;
Best Local Similarity 100.0%; Pred. No. 4.4;

```

```
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 299 gaagactatctcaagcaagc 319
Db 116245 GAAGACTATCTCAAGCAAGC 116265

RESULT 14
AL611984/c 206591 bp DNA linear HTG 13-OCT-2001
LOCUS Mus musculus chromosome 4 clone RP23-41608, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL611984
VERSION AL611984.6 GI:16151499
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 206591)
REFERENCE Direct Submission
AUTHORS Submitted (12-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
TITLE CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk
JOURNAL requests: clonerequest@sanger.ac.uk
On Oct 15, 2001 this sequence version replaced gi:16116600.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM416J8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 203072 bases at least Q40
Consensus quality: 203785 bases at least Q30
Consensus quality: 204261 bases at least Q20
Insert size: 204991; sum-of-contigs
Insert size: 196126; 3.5% error; agarose-fp
Quality coverage: 9.25x in Q20 bases; sum-of-contigs Quality
coverage: 9.86x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
Source location/Qualifiers
1..206591
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-416J8"
/clone_lib="PC1-23"
1..12340
/note="assembly_fragment:01861
fragment_chain:1"
misc_feature 14634..14533
/note="assembly_fragment:00110
fragment_chain:1"
misc_feature 14634..25420
/note="assembly_fragment:02706
fragment_chain:1"
misc_feature 25521..38180
/note="assembly_fragment:02641
fragment_chain:1"
misc_feature 38281..45684
/note="assembly_fragment:01195
fragment_chain:1"
misc_feature 45785..49116
/note="assembly_fragment:00318
fragment_chain:1"
misc_feature 49217..103493
/note="assembly_fragment:01418
fragment_chain:1"
misc_feature 103594..122215
/note="assembly_fragment:04869
fragment_chain:1"
misc_feature 122316..125353
/note="assembly_fragment:00804
fragment_chain:1"
misc_feature 125454..127674
/note="assembly_fragment:02278
fragment_chain:1"
misc_feature 127775..132117
/note="assembly_fragment:01785
fragment_chain:2"
misc_feature 132218..139371
/note="assembly_fragment:03689
fragment_chain:2"
misc_feature 139472..155577
/note="assembly_fragment:02580
fragment_chain:3"
misc_feature 15678..188900
/note="assembly_fragment:04111
fragment_chain:3"
misc_feature 189001..195856
/note="assembly_fragment:00631
fragment_chain:4"
misc_feature 195957..198542
/note="assembly_fragment:03866
fragment_chain:4"
misc_feature 198643..206591
/note="assembly_fragment:02750
fragment_chain:4"
clone_end:T7
vector_side:right"
BASE COUNT 53779 a 47056 c 47603 g 56551 t 1602 others
ORIGIN
Query Match 1.1%; Score 21; DB 2; Length 206591;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1044 cccgcgatatctcttaaaa 1064
Db 156967 CCCTGCCGATATCTCTAAAA 156947

RESULT 15
AL645546 225609 bp DNA linear HTG 13-NOV-2001
LOCUS Mus musculus chromosome 13 clone RP23-124G18, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL645546
VERSION AL645546.4 GI:16944328
KEYWORDS HTG: HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
REFERENCE Direct Submission
AUTHORS Submitted (08-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
TITLE Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
JOURNAL humquerry@sanger.ac.uk
COMMENT On Nov 15, 2001 this sequence version replaced gi:16754763.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
```

```

----- Project Information
Center project name: bml24G18
----- Summary Statistics
Assembly program: XGAP; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 224286 bases at least Q40
Consensus quality: 22463 bases at least Q30
Consensus quality: 22493 bases at least Q20
Insert size: 225409; sum-of-contigs
Insert size: 216124; 7.8% error; agarose-fp
Quality coverage: 11.91x in Q20 bases; sum-of-contigs Quality
coverage: 13.08x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence..
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
    Location/Qualifiers
        1..225609
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="13"
            /clone="RP23-124G18"
            /clone_1fb="RPC1-23"
            1..101657
            /note="assembly_fragment:01246"
            clone_end:T7
            vector_side:left"
misc_feature
    101758..190829
    /note="assembly_fragment:02144.0"
    190930..225609
    /note="assembly_fragment:03429"
    clone_end:SP6
    vector_side:right"
BASE COUNT 66198 a 45409 c 44739 g 69047 t 216 others
ORIGIN
Query Match 1.1%; Score 21; DB 2; Length 225609;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 acacagcagccttagcacca 238
|||||
Db 147727 ACACAGCAGCCTTAGCACCA 147747

```

Search completed: August 20, 2002, 03:20:54
 Job time: 7604 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:12:06 ; Search time 28.88 Seconds

(without alignments)
2184.554 Million cell updates/sec

Title: US-09-662-812-2

Perfect score: 568
Sequence: 1 MGLFHLTLFGLLCSLPISL.....TEGDAKNPVLAGLIKIKIVE 568

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_032802:*

1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568	100.0	568	22	AAV72972
2	327	57.6	348	20	AAV34620
3	241	42.4	245	20	AAV34619
4	12	2.1	468	20	AAV36830
5	10	1.8	435	22	AAU35103
6	8	1.4	52	22	AAV66330
7	8	1.4	77	22	AAV99877
8	8	1.4	94	22	AAV80047
9	8	1.4	104	22	AAV29407
10	8	1.4	109	22	AAU37519
11	8	1.4	142	21	AAV41090

12	8	1.4	207	22	ABG06236	Novel human diagno
13	8	1.4	254	22	ABG14349	Novel human diagno
14	8	1.4	372	20	AAW89449	A gida2 polypeptid
15	8	1.4	379	21	AAV10364	Arabidopsis thalia
16	8	1.4	382	21	AAV10363	Arabidopsis thalia
17	8	1.4	401	22	AAV38082	Streptococcus pneu
18	8	1.4	411	16	AAV67733	Artematic dihydridol
19	8	1.4	433	22	AAV34397	Staphylococcus aur
20	8	1.4	435	20	AAW89448	A gida2 polypeptid
21	8	1.4	435	22	AAV37266	Staphylococcus aur
22	8	1.4	444	20	AAW89450	A gida2 polypeptid
23	8	1.4	444	22	AAV37761	Streptococcus pneu
24	8	1.4	447	22	AAV98357	Escherichia coli p
25	8	1.4	448	22	AAV90237	C glutamicum prote
26	8	1.4	448	22	AAV80046	Corynebacterium g1
27	8	1.4	498	22	AAV03145	Streptococcus pyog
28	8	1.4	520	22	AAV41718	Propionibacterium
29	8	1.4	523	21	AAV09513	Arabidopsis thalia
30	8	1.4	523	21	AAV38975	Arabidopsis thalia
31	8	1.4	528	21	AAV09512	Arabidopsis thalia
32	8	1.4	558	21	AAV38974	Arabidopsis thalia
33	8	1.4	568	21	AAV15185	Partial soybean ze
34	8	1.4	568	21	AAV15186	Wheat zeta-caroten
35	8	1.4	640	22	ABG03549	Novel human diagno
36	7	1.2	57	22	AAV44473	Thrombin immunogen
37	7	1.2	34	21	AAV28021	Human secreted pro
38	7	1.2	52	22	ABV44489	Novel human diagno
39	7	1.2	58	22	ABV44489	Peptide #11995 enc
40	7	1.2	58	22	AAV65610	Human brain expres
41	7	1.2	58	22	AAV78237	Human bone marrow
42	7	1.2	58	22	AAV38500	Peptide #12537 enc
43	7	1.2	61	21	AAV56502	Human prostate can
44	7	1.2	69	19	AAV70258	Leishmania antigen
45	7	1.2	73	22	ABV68835	Drosophila melanog

ALIGNMENTS

RESULT 1

AAV72972 standard; Protein: 568 AA.

XX AAV72972:

XX 13-JUN-2001 (first entry)

DE Chlamydia pneumoniae outer membrane protein (OMP).

XX Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;

KW antibody; medicament; Chlamydia infection.

OS Chlamydia pneumoniae.

XX

XX WO200121804-A1.

XX 29-MAR-2001.

PD

XX 15-SEP-2000; 2000WO-CA01088.

PF

XX 20-SEP-1999; 99US-0154652.

PR

XX (AVET) AVENTIS PASTEUR LTD.

PA

XX Murdin AD, Oomen RP, Wang J, Dunn P;

PI WPI; 2001-244939/25.

DR N-PSDB; AAD03024.

XX

XX Novel Chlamydia pneumoniae outer membrane protein and polynucleotides encoding them, useful as components of vaccines for treating Chlamydia infections, and for detecting Chlamydia infections in the body fluids of mammals -

AAV34619
 ID AAV34619 standard; Protein: 245 AA.
 XX
 AC AAV34619;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Chlamydia pneumoniae transmembrane protein sequence.
 XX
 KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope.
 XX
 OS Chlamydia pneumoniae.
 PN WO927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI; 1999-357842/30.
 XX
 PS Genome sequence of Chlamydia pneumoniae
 XX
 PS Page 639-640; Disclosure: 1912pp; English.
 XX
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 CC
 XX
 SQ Sequence 245 AA;
 XX

Query Match 42.4%; Score 241; DB 20; Length 245;
 Best Local Similarity 100.0%; Pred. No. 2.3e-220;
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 328 KLVNALALDELGCDTPKLEYTERLVQPHYNETLALFSKGRQLQWKRNVIIIVPDDPQ 387
 DB 5 klvnaiallalgcdtpkleyterlvqphynetlalfskgrlqwmkrnviiivpddpq 64

OY 388 EREBLSTGTGLEQILTFELPRLPKKAYLPCIVYLLASQKQLATTAISPLSHSHQBAL 447
 DB 65 ereblstgtgleqiltfelrplpkkaylpcivyllasqkqlattaisfshshebal 124

OY 448 DLLFOAAKLPGEPITRAVADLAIVNLKDEPKRSLHDYAKKLIOETLLFVDTENORPHR 507
 DB 125 dllfqaaklpgepitrayadalaivnlkdepkrsldyakkliqetllfvdtengrphr 184

OY 508 SMPYLKRVQVPESTRKMLDLLETLATSKSSEDIRLLIQLTGSDARNFPVLAQLIKIV 567
 DB 185 smpylrvqvpestrkmlldletlatksseidirlliqltgsgdarnfpvlaglikiv 244

OY 568 E 568
 DB 245 e 245

RESULT 4
 AAY36830
 ID AAY36830 standard; Protein: 468 AA.
 XX
 AC AAY36830;
 XX
 DT 07-OCT-1999 (first entry)
 XX
 DE Chlamydia trachomatis lipoprotein sequence.
 XX
 KM Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KM paratrachoma; inclusion conjunctivitis; genital disease; perinephalitis;
 KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KM Bartholinitis; pneumopathy; venereal lymphogranulomatosis.
 XX
 OS Chlamydia trachomatis.
 PN WO928475-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 27-NOV-1998; 98WO-IB01939.
 XX
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI; 1999-371125/31.
 XX
 PS Genome sequence of Chlamydia trachomatis
 XX
 PS Disclosure: Page 708-710; 1755pp; English.
 XX
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nongonococcal trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perinephalitis, Bartholinitis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 CC
 XX
 SQ Sequence 468 AA;
 XX

Query Match 2.1%; Score 12; DB 20; Length 468;
 Best Local Similarity 100.0%; Pred. No. 0.012;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 141 PVIRLEAAYRLA 152
 DB 142 pvirleaayrla 153

RESULT 5
 AAU35103
 ID AAU35103 standard; Protein: 435 AA.
 XX
 AC AAU35103;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Enterococcus faecalis cellular proliferation protein #390.
 XX
 KM Antisense: prokaryotic cellular proliferation protein;

KM		antibiotic; antibacterial; drug design.
XX		
OS		Enterococcus faecalis.
XX		
PN		W0200170955-A2.
XX		
PD		27-SEP-2001.
XX		
PF		21-MAR-2001; 2001WO-US09180.
XX		
PR		21-MAR-2000; 2000US-191078P.
PR		23-MAY-2000; 2000US-206848P.
PR		26-MAY-2000; 2000US-207727P.
PR		23-OCT-2000; 2000US-242578P.
PR		27-NOV-2000; 2000US-253625P.
PR		22-DEC-2000; 2000US-257931P.
PR		16-FEB-2001; 2001US-269308P.
XX		
PA		(ELIT-) ELITRA PHARM INC.
PI		
PI		Haselbeck R, Ohlsen KI, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI		Yamamoto RT, Xu HH;
XX		
DR		WPI: 2001-611495/70.
DR		N-PDSB: AAS52962.
XX		
PT		New polynucleotides for the identification and development of
PT		antibiotics, comprise sequences of antisense nucleic acids -
XX		
PS		Example 3; Seq ID No 10696; 511pp; English.
XX		
CC		The invention relates to antisense inhibitors of genes essential to
CC		prokaryotic cellular proliferation, their use in identifying the
CC		genes, their use in the discovery of novel antibiotics, the essential
CC		genes themselves and the encoded proteins. The prokaryotes used are
CC		Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC		pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC		invention is also useful for the identification of potential new targets
CC		for antibiotic development. The antisense nucleic acids can also be used
CC		to identify proteins used in proliferation, to express these proteins,
CC		and to obtain antibodies capable of binding to the expressed proteins.
CC		The proteins can be used to screen compounds in rational drug discovery
CC		programmes. The antisense nucleic acid sequence is also useful to screen
CC		for homologous nucleic acids which are required for cell proliferation in
CC		a wide variety of organisms. The present sequence represents an
CC		essential prokaryotic cellular proliferation protein.
CC		Note: The sequence data for this patent did not form part
CC		of the printed specification, but was obtained in electronic
CC		format directly from WIPO at
CC		ftp.wipo.int/pub/published_pct_sequences.
XX		
SQ		Sequence 435 AA;
XX		
OY	85 TTGGAGGAGS 94	
DB		
	6 tllggaglegs 15	
XX		
RESULT	6	
AAB66330		
ID	AAB66330 standard; Protein: 52 AA.	
XX		
AC	AAB66330;	
DT	05-APR-2001 (first entry)	
DE	Synechocystis zeta-caroteiin desaturase SEQ ID NO: 78.	
XX		

```

KW Linolate isomerase; conjugated linoleic acid; conjugated linolenic acid
KM CLA; oil; lipase-like protein; acetyltransferase-like protein.
XX Synecocystis sp.
XX WO200100846-A2.
XX PD 04-JAN-2001.
XX PF 30-JUN-2000; 2000WO-US18149.
XX PR 30-JUN-1999; 99US-0141798.
XX PR 28-APR-2000; 2000US-0561077.
XX PA (DCVB-) DCV INC DBA BIO-TECH RESOURCES.
XX PI Rosson RA, Deng M, Grund AD, Peng SS;
XX WPI: 2001-102894/11.
XX DR
XX PT Isolated linoleate isomerase proteins and nucleic acids, useful in
XX methods for the production of conjugated linoleic acid or conjugated
XX linolenic acid (CLA) -
XX PS Example 13; Page 242; 244pp; English.
XX CC The present invention provides the protein and coding sequences for the
XX CC linolate isomerase enzyme from Propionibacterium acnes. This enzyme is
XX CC useful for producing conjugated linoleic and linolenic acids (CLAs) from
XX CC linoleic and linolenic acids in oils. Also provided are the sequences of
XX CC an acetyltransferase-like protein and a lipid-like protein from the same
XX CC organism.
XX SQ Sequence 52 AA;
OY 86 IIGAGLAG 93
   |||||
Db 4 IIGAGLAG 11

RESULT 7
ID AAB99877 standard; Protein; 77 AA.
AC AAB99877;
DT 21-SEP-2001 (first entry)
XX Physcomitrella patens 42_ck10_g09fwd protein.
DE
XX Tocopherol and carotenoid metabolism related protein; TCMRP; synthesis;
XX Physcomitrella patens; moss; algae; microorganism; fungus; plant;
KW identification; genome mapping; modulation; evolutionary study;
KW cellular production; fine chemical.
XX Physcomitrella patens.
OS WO200144276-A2.
PN
XX 21-JUN-2001.
PD
XX 14-DEC-2000; 2000WO-EPI2698.
PF
XX 16-DEC-1999; 99US-0171121.
PR
XX (BADI ) BASF PLANT SCI GMBH.
PA
XX Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;

```

PI Frank M, Freund A, Duwenig E, Schmidt R, Reeki R, Badur R;
XX WPI: 2001-398121/42.
DR N-PSDB; AAH44250.
XX
PT Tocopherol and carotenoid metabolism related protein (TCMRP), used to
PT produce fine chemicals, is isolated from mosses, algae, microorganisms,
PT fungi, plants, or their fragments -
XX
PS Claim 28; Page 119; 123pp; English.
XX
CC The present invention describes isolated tocopherol and carotenoid
CC metabolism related proteins (TCMRP) (I) from mosses or algae,
CC microorganisms or fungi, plants, or its fragments. (I) can be used as
CC enzymes in the production of fine chemicals or in the metabolism of
CC tocopherol and carotenoids. (I) also assist in transmembrane transport.
CC The fine chemicals that can be produced include lipids, fatty acids,
CC vitamins, cofactors, enzymes, amino acids, and nucleotide bases.
CC Nucleotide sequences, proteins, vectors and host cells from the present
CC invention can be used: (a) to identify mosses related to Physcomitrella
CC patens; (b) in mapping genomes of mosses related to Physcomitrella
CC studies; (c) in the modulation of TCMRP activity; (d) in evolutionary
CC in the determination of functional TCMRP regions; (f) and
CC encode the Physcomitrella patens TCMRP proteins given in AAH44262
CC AAH9889. AAH44212 to AAH44221 represent nucleotide sequence used in
CC the exemplification of the present invention.
XX
SQ Sequence 77 AA:

Query Match 1.4%; Score 8; DB 22; Length 77;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 ITGAGLAG 93
|||
Db 49 IIGAGIAG 56

RESULT 8
AAB80047
ID AAB80047 standard; Protein; 94 AA.
XX
AC AAB80047;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:828.
XX
KM Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KM fine chemical production; microorganism; organic acid; nucleoside;
KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KM lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KM carbohydrate; aromatic compound; cofactor; polypeptide; enzyme.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100843-A2.
PD
PD 04-JAN-2001.
XX
XX 23-JUN-2000; 2000MO-IB00923.
PF
XX 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
XX (BADI) BASF AG.
XX
XX Pompejus M, Kroegeer B, Schroeder H, Zeider O, Haberhauer G;
PI WPI: 2001-137957/14.
DR N-PSDB; AAF72166.
XX
PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
PS Claim 20; Page 1301-1302; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polypeptides and enzymes.
XX
SQ Sequence 94 AA;

Query Match 1.4%; Score 8; DB 22; Length 94;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 86 IIGAGLAG 93
|||||||
Db 5 IIGAGLAG 12

RESULT 9
ABG29407
ID ABG29407 standard; Protein: 104 AA.
XX
XX
AC ABG29407;
XX
XX 18-FEB-2002 (first entry)
DT
XX
DE Novel human diagnostic protein #29398.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB: AAS93594.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID NO 59766; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II) (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 104 AA;
SQ

Query Match 1.4%; Score 8; DB 22; Length 104;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 88 GAGLAGSS 95
|||||||
Db 83 GAGLAGSS 90

RESULT 10
AAU37519
ID AAU37519 standard; Protein: 109 AA.
XX
XX
AC AAU37519;
XX
XX 14-FEB-2002 (first entry)
DT
XX
DE Staphylococcus aureus cellular proliferation protein #1689.
XX
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
OS
XX WO200170955-A2.
PN
XX 27-SEP-2001.
PD
XX 21-MAR-2001; 2001WO-US09180.
PF
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
PI Haselbeck R, Ohlsen KM, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR N-PSDB: AAS55378.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PT
XX Example 3; Seq ID No 13112; 51pp; English.
PS
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 109 AA;
SQ

Query Match 1.4%; Score 8; DB 22; Length 109;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 IGAGLAGS 94
|||||||
Db 8 igaglags 15

RESULT 11
AAB41090
ID AAB41090 standard; Protein; 142 AA.
XX
AC AAB41090;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human OREFX OREF854 polypeptide sequence SEQ ID NO:1708.
XX
XX Human: open reading frame; OREFX detection; cytostatic; hepatotropic;
KM vulnerable; antiparietic; antiparkinsonian; nootropic; neuroprotective;
KM anticonvulsant; osteopathic; antiallergic; immunosuppressant; cardiant;
KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KM hypotensive; dermatological; immunosuppressive; antinflammatory;
KM antiviral; antibacterial; antifungal; antihemmatic; antihypoid;
KM antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KM cholesterol ester storage; systemic lupus erythematosus; infection;
KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KM bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;
KM thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
XX
PR 02-APR-1999; 99US-0127636.
XX
PR 05-APR-1999; 99US-0127728.
XX
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
XX
DR N-PSDB; AAC75299.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 1359; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
which represent the human OREFX open reading frames 1 to 3161. The OREFX
sequences have activities such as: cytostatic; hepatotropic; vulnerable;
antiparietic; antiparkinsonian; nootropic; neuroprotective;
osteopathic; anticonvulsant; antiallergic; immunosuppressive;
immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
antihypoid; and antianaemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an OREFX-associated disorder. The

CC nucleic acids can be used to express OREFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 142 AA:

Query Match 1.4%; Score 8; DB 21; Length 142;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAG 93
|||||||
Db 28 iigaglag 35

RESULT 12
ABG06236
ID ABG06236 standard; Protein; 207 AA.
XX
AC ABG06236;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6227.
XX
KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS70423.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 36595; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 207 AA;

Query Match 1.4%; Score 8; DB 22; Length 207;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 88 GAGLAGSS 95
|||||
Db 86 gaglagss 93

RESULT 13
ABG14349
ID ABG14349 standard; Protein: 254 AA.
XX
AC ABG14349;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #14340.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS78536.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 44708; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 254 AA;

Query Match 1.4%; Score 8; DB 22; Length 254;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 88 GAGLAGSS 95
|||||
Db 56 gaglagss 63

RESULT 14
AAM89449
ID AAM89449 standard; Protein: 372 AA.
XX
AC AAM89449;
XX
DT 18-MAR-1999 (first entry)
XX
DE A gida2 polypeptide fragment.
XX
KW gida2; Staphylococcus aureus WCUH29; bacterial infection;
KW Helicobacter pylori infection; cancer; ulcer; gastritis; antibacterial;
KW wound treatment; bacterial adhesion; matrix protein.
XX
OS Staphylococcus aureus.
XX
PN EP889131-A2.
XX
PD 07-JAN-1999.
XX
PF 30-JUN-1998; 98EP-0305203.
XX
PR 01-JUL-1997; 97US-0051380.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Burnham M, Debouck CM, Kallender H, Lenox AL, Mooney JL;
PI Palmer LM, Zhong YF;
XX
DR WPI: 1999-062662/06.
DR N-PSDB; AAV82085.
XX
PT New isolated gida2 polypeptide from Staphylococcus aureus - used to
PT diagnose, treat and prevent bacterial infections e.g. S. aureus and
PT H. pylori, related cancers, ulcers and gastritis and to prevent
PT adhesion of bacteria to matrix proteins
XX
PS Claim 1; Page 6; 41pp; English.

CC The present sequence represents a gida2 protein fragment of
CC Staphylococcus aureus WCUH29. The gida2 proteins and nucleic acids
CC are used to treat conditions requiring increased activity or expression
CC of gida2, while conditions (e.g. bacterial infections) requiring
CC inhibition of such activity or expression are treated by administering
CC an antagonist, inhibitory nucleic acid or competitive polypeptide.
CC Infection by S. aureus is treated, but also Helicobacter pylori
CC infections and related cancers, ulcers and gastritis. These antibacterial
CC agents may also be used to treat in-dwelling devices to prevent infection
CC or generally as wound treatments to prevent adhesion of bacteria to
CC matrix proteins.

PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

OY 86 IIGAGLAG 93
Db 51 IIGAGLAG 58

Search completed: August 20, 2002, 04:25:18
Job time: 792 sec

Query Match 1.4%; Score 8; DB 21; Length 379;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compgen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 03:12:10 ; Search time 58.23 Seconds

(without alignments)
1083.461 Million cell updates/sec

Title: US-09-662-812-2

Perfect score: 2833
Sequence: 1 MGFLHFLFLGLLCSLPISL.....TEGDKNFPVLAGLLIKIVE 568

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2833	100.0	568	22	AAV72972	Chlamydia pneumoni
2	1619	57.1	348	20	AAV34620	Chlamydia pneumoni
3	1215	42.9	245	20	AAV34619	Chlamydia pneumoni
4	1092	38.5	468	20	AAV36830	Chlamydia trachoma
5	296	10.4	96	20	AAV37834	Amino acid sequenc
6	137.5	4.9	1015	22	ABG04636	Novel human diagno
7	132.5	4.7	514	22	ABR64398	Drosophila melanog
8	129.5	4.6	7201	22	ABR71136	Drosophila melanog
9	128.5	4.5	1870	22	ABG21017	Novel human diagno
10	127.5	4.5	2017	22	ABG06301	Novel human diagno
11	126	4.4	631	17	AAV9675	RHAMM 1-2a isoform

12	123.5	4.4	606	17	AAV9673	Receptor for hyalu
13	123.5	4.4	630	18	AAW39166	Mouse RHAMM protei
14	121.5	4.3	1838	22	AAU00015	Human Plexin-B2.
15	121.5	4.3	1849	22	ABH11818	Human plexin-B1/SE
16	121.5	4.3	2985	19	AAW56643	DNA-dependent prot
17	121.5	4.3	3360	19	AAW56642	Protein Kinase cat
18	119.5	4.2	688	22	AAV64567	Human vacuole prot
19	118.5	4.2	716	22	AAW93421	Human polyptide,
20	118.5	4.2	2117	22	ABR59379	Drosophila melanog
21	118.5	4.2	2117	22	ABR67222	Drosophila melanog
22	117.5	4.1	1111	22	ABG09489	Novel human diagno
23	117.5	4.1	1336	22	AAV60489	Human cell cycle a
24	117.5	4.1	2688	22	AAW40883	Human polyptide
25	116.5	4.1	740	22	ABR60942	Drosophila melanog
26	115.5	4.1	1111	22	AAW80108	Human protein SEQ
27	115.5	4.1	1111	22	AAW80109	Human protein SEQ
28	115.5	4.1	1569	22	ABG15232	Novel human diagno
29	115.5	4.1	1743	19	AAW98879	H. pylori GHPD 175
30	115	4.1	1185	19	AAW98508	H. pylori GHPD 123
31	114.5	4.0	883	22	ABR59757	Drosophila melanog
32	114	4.0	746	21	AAV42530	Arabidopsis thalia
33	114	4.0	840	21	AAV42529	Arabidopsis thalia
34	114	4.0	932	21	AAV42528	Arabidopsis thalia
35	114	4.0	1179	22	AAW01107	CBE 110 protein se
36	114	4.0	1217	22	AAW52358	Putative TBP inter
37	114	4.0	1388	19	AAW56475	Protein with Rho p
38	114	4.0	2076	22	AAU34319	Staphylococcus aur
39	114	4.0	2186	22	AAU37320	Staphylococcus aur
40	114	4.0	2663	22	AAW39097	Human polyptide
41	113.5	4.0	1119	22	AAW79124	Human protein SEQ
42	113.5	4.0	2207	22	AAU32041	Novel human secret
43	113	4.0	1179	22	AAU37669	Streptococcus pneu
44	112.5	4.0	1690	22	ABR61144	Drosophila melanog
45	112.5	4.0	1690	22	ABR61173	Drosophila melanog

ALIGNMENTS

RESULT 1
AAV72972 standard: Protein; 568 AA.
ID AAV72972:
AC AAV72972:
DT 13-JUN-2001 (first entry)
DE Chlamydia pneumoniae outer membrane protein (OMP).
KW Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;
KW antibody; medicament; Chlamydia infection.
XX OS Chlamydia pneumoniae.
XX PN WO200121804-A1.
XX PD 29-MAR-2001.
XX PF 15-SEP-2000; 2000MC-CA01088.
XX PR 20-SEP-1999; 99US-0154652.
XX PA (AVER) AVENTIS PASTEUR LTD.
XX PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX DR WPI: 2001-244939/25.
XX DR N-PSDB: AAD03024.
XX PT Novel Chlamydia pneumoniae outer membrane protein and polynucleotides
XX PT encoding them, useful as components of vaccines for treating Chlamydia
XX PT infections, and for detecting Chlamydia infections in the body fluids
XX PT of mammals -

XX Claim 16; Fig 1; 82pp; English.

CC The present sequence is Chlamydia pneumoniae OMP (outer membrane
CC protein). OMP is useful in the production of vaccines, antibodies
CC and pharmaceutical compositions which are useful for treating or
CC preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or
CC C. pecorum) infections. They are also useful as diagnostic reagents for
CC detecting Chlamydia infection which involves assaying a body fluid
CC of a mammal to be tested for the components. The OMP vaccine is
CC useful in the preparation of a medicament for preventing and/or
CC treating Chlamydia infection. The primers derived from OMP gene are
CC also useful for detecting and/or identifying Chlamydia in a biological
CC material. OMP antibodies are also useful as reagents for purifying OMP
CC from a biological sample which involves carrying out antibody-based
CC affinity chromatography with the biological sample. OMP gene is
CC also useful in gene therapy.

CC Sequence 568 AA;

Query Match 100.0%; Score 2833; DB 22; Length 568;

Best Local Similarity 100.0%; Pred. No. 1,1e-227; Mismatches 0; Indels 0; Gaps 0;

Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLFHLTLFGLLCSLPISLVAKPEESVGHKILYISTOSTOQALATYLEALDAYGDHDF 60
DB 1 mglfhltlfllgllcslpislvakfpesvghkilyistgstqalatyaleadaygdhdf 60
QY 61 VLKRGEDYLKQSIHSSDPQTRKSTITIGAGLSSEALDVLVSQAMETADPLQQLVLSAV 120
DB 61 vlkrigedylkqsihssdpqtrkstltaglsgsealdvlsqametaadplqqlvlvsav 120
QY 121 SGHGLKTSDDLFFKALASPYVIRLEAAYRLANKTKFVIDHLSFTHKLPREIQCUSA 180
DB 121 sghglktsddllfkalspyvirleaaayrlanktkfvldhlsfthklpreiqcisa 180
QY 121 sghglktsddllfkalspyvirleaaayrlanktkfvldhlsfthklpreiqcisa 180
DB 121 sghglktsddllfkalspyvirleaaayrlanktkfvldhlsfthklpreiqcisa 180
QY 181 IFRLTEESDAYIRDLAANKSAIRSATALQIGYQOKRFLPLRLNLTSSASPODEAI 240
DB 181 ifrlteesdayirldlaanksairsatalqigyqokrflplrlnltsaspoadeai 240
QY 181 ifrlteesdayirldlaanksairsatalqigyqokrflplrlnltsaspoadeai 240
DB 181 ifrlteesdayirldlaanksairsatalqigyqokrflplrlnltsaspoadeai 240
QY 241 LYALGKLDGSGSYNIRKQLOKPPVDVTLAAQAALIALGKEEDALPVTKQALEERPAL 300
DB 241 lyalgkldgsgsynirkqlokpvdvltlaaqaalialgkeedalpvtkqaleerpal 300
QY 241 lyalgkldgsgsynirkqlokpvdvltlaaqaalialgkeedalpvtkqaleerpal 300
DB 241 lyalgkldgsgsynirkqlokpvdvltlaaqaalialgkeedalpvtkqaleerpal 300
QY 301 YALRHLPSEIGIPALPIFLTKNSEAKLVNALLELGCDFPKLXETTERLVQPHNE 360
DB 301 yalrhlpseigipalpifltknseaklvnalallelgcdfpklexetterlvqphne 360
QY 301 yalrhlpseigipalpifltknseaklvnalallelgcdfpklexetterlvqphne 360
DB 301 yalrhlpseigipalpifltknseaklvnalallelgcdfpklexetterlvqphne 360
QY 361 TLALSFSGRTLOMKRVNITVPDPOERERLSTTRGLEBOILFTFLRLEKAVLPCTY 420
DB 361 tlaifsfsgrtlqmkrvnitvpdpoeerlstrgleboilftflrlekeavlpcty 420
QY 361 tlaifsfsgrtlqmkrvnitvpdpoeerlstrgleboilftflrlekeavlpcty 420
DB 361 tlaifsfsgrtlqmkrvnitvpdpoeerlstrgleboilftflrlekeavlpcty 420
QY 421 KLASQRTQALATTAISPLSTSHOBALDLFOAAKLEGEPIIRAYADLAIVNLTDPKPK 480
DB 421 klsaqrtqalattaisplstshobaldlfoaaklegepiirayadlaivnltdpkpk 480
QY 421 klsaqrtqalattaisplstshobaldlfoaaklegepiirayadlaivnltdpkpk 480
DB 421 klsaqrtqalattaisplstshobaldlfoaaklegepiirayadlaivnltdpkpk 480
QY 481 RSLHDYKKKLIOETFLFDFENORPHSMPLYAOVPESTKMLDITELTSLASSED 540
DB 481 rslhdyykkklifoetflfddfennorphsmplyaovpestkmlditelstlasksed 540
QY 481 rslhdyykkklifoetflfddfennorphsmplyaovpestkmlditelstlasksed 540
DB 481 rslhdyykkklifoetflfddfennorphsmplyaovpestkmlditelstlasksed 540
QY 541 IRLIQLMTEGDANKFPYLGLIKRIIVE 568
DB 541 irllqlmtegdankfpylglilkrilive 568

RESULT 2

AAV34620 ID AAV34620 standard; Protein; 348 AA.

XX AAV34620;

DT 13-SEP-1999 (first entry)

XX Chlamydia pneumoniae lipoprotein sequence.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN W09927105-A2.

XX 03-JUN-1999.

PD 20-NOV-1998; 98WO-1B01890.

PF 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

PA Griffais R;

DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

PS Page 640; Disclosure: 1912pp; English.

CC AAV34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.

XX Sequence 348 AA;

Query Match 57.1%; Score 1619; DB 20; Length 348;

Best Local Similarity 100.0%; Pred. No. 1e-126; Mismatches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLFHLTLFGLLCSLPISLVAKPEESVGHKILYISTOSTOQALATYLEALDAYGDHDF 60
DB 1 mglfhltlfllgllcslpislvakfpesvghkilyistgstqalatyaleadaygdhdf 60
QY 21 mglfhltlfllgllcslpislvakfpesvghkilyistgstqalatyaleadaygdhdf 80
DB 21 mglfhltlfllgllcslpislvakfpesvghkilyistgstqalatyaleadaygdhdf 80
QY 61 VLKRGEDYLKQSIHSSDPQTRKSTITIGAGLSSEALDVLVSQAMETADPLQQLVLSAV 120
DB 61 vlkrigedylkqsihssdpqtrkstltaglsgsealdvlsqametaadplqqlvlvsav 140
QY 81 vlkrigedylkqsihssdpqtrkstltaglsgsealdvlsqametaadplqqlvlvsav 140
DB 81 vlkrigedylkqsihssdpqtrkstltaglsgsealdvlsqametaadplqqlvlvsav 140
QY 121 SGHGLKTSDDLFFKALASPYVIRLEAAYRLANKTKFVIDHLSFTHKLPREIQCUSA 180
DB 121 sghglktsddllfkalspyvirleaaayrlanktkfvldhlsfthklpreiqcusa 200
QY 141 sghglktsddllfkalspyvirleaaayrlanktkfvldhlsfthklpreiqcisa 200
DB 141 sghglktsddllfkalspyvirleaaayrlanktkfvldhlsfthklpreiqcisa 200
QY 181 IFRLTEESDAYIRDLAANKSAIRSATALQIGYQOKRFLPLRLNLTSSASPODEAI 240
DB 181 ifrlteesdayirldlaanksairsatalqigyqokrflplrlnltsaspoadeai 260
QY 201 ifrlteesdayirldlaanksairsatalqigyqokrflplrlnltsaspoadeai 260
DB 201 ifrlteesdayirldlaanksairsatalqigyqokrflplrlnltsaspoadeai 260
QY 241 LYALGKLDGSGSYNIRKQLOKPPVDVTLAAQAALIALGKEEDALPVTKQALEERPAL 300
DB 241 lyalgkldgsgsynirkqlokpvdvltlaaqaalialgkeedalpvtkqaleerpal 320
QY 261 lyalgkldgsgsynirkqlokpvdvltlaaqaalialgkeedalpvtkqaleerpal 320
DB 261 lyalgkldgsgsynirkqlokpvdvltlaaqaalialgkeedalpvtkqaleerpal 320
QY 301 YALRHLPSEIGIPALPIFLTKNSEA 327
DB 301 yalrhlpseigipalpifltknsea 347

RESULT 3

[illegible]

Result	4
AA36830	standard; Protein; 468 AA.
AA36830	standard; Protein; 468 AA.
AA36830	standard; Protein; 468 AA.
AA36830	standard; Protein; 468 AA.
07-OCT-1999	(first entry)
Chlamydia trachomatis	lipoprotein sequence.
Vaccine; eye disease; conventional trachoma; nonendemic trachoma; paratrachoma; inclusion conjunctivitis; genital disease; perithenitis; nongonococcal urethritis; epididymitis; cervicitis; salpingitis; Bartholinitis; pneumopathy; venereal lymphogranulomatosis.	
Chlamydia trachomatis.	
MO9928475-A2.	
10-JUN-1999.	
27-NOV-1998;	98WO-IB01939.
04-NOV-1998;	98US-0107077.
28-NOV-1997;	97FR-0015041.
17-DEC-1997;	97FR-0016034.
(GEST) GENSET.	
Griffais R;	
WPI; 1999-371125/31.	
Genome sequence of Chlamydia trachomatis	
Disclosure; Page 708-710; 1755pp; English.	
AA36754-X37949 are encoded by open reading frames (ORFs) of the genome of Chlamydia trachomatis (see AA201425). The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perithenitis, Bartholinitis; pneumopathy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.	
Sequence 468 AA:	
Query Match	38.5%; Score 1092; DB 20; Length 468;
Best Local Similarity	51.5%; Pred. No. 1.5e-82;
Matches 238; Conservative 67; Mismatches 155; Indels 2; Gaps 1;	
1 MGLFLLFGLLCSLPISLVAKFPESVGHKILYISTOSTQDALTYEALDAYGDHDF	60
4 mglslrlafgfl-sftlsascdffpsvsqrllfscqksvpqaleayleasvtyqhdts	61
61 VLARKIGEDYLKSHSSDPQTRKSTIIIGAGLGSSEALDVLQAMETADPLQOLIVLSAV	120
62 vlrkigedyLkshssdpqtrkStIIIGAGLGSSEALDVLQAMETADPLQOLIVLSAV	121
121 SGLHKTSDDLFLKALASBPVIRLEAAVRLANLNTKVIDHLSFIRHKLPEDIOCLSA	180
122 tsqLktsdskllfkgltsashpvrleaaYrLacmknksvdylysfyiklpeiqnlaet	181
181 IFLRTEESDAYIRIDLAAKSAIRSATALDIGEYQQRFLPTLRNLNTSAPDQDAI	240
182 iflRteesdayIRIDLAAKSAIRSATALDIGEYQQRFLPTLRNLNTSAPDQDAI	241

OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US092231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL15239.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 PS Disclosure; SEQ ID NO 40200; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 CC
 CC
 CC Sequence 7201 AA;
 SO

Query Match 4.6%; Score 129.5; DB 22; Length 7201;
 Best Local Similarity 22.2%; Pred. No. 1.5;
 Matches 135; Conservative 78; Mismatches 217; Indels 177; Gaps 28;

QY 41 QGALATYIEALDAGDHFVLRKIGEDYTKQSIHS-SDPQTRKSTTIGAGLAGSSEALD 99
 DB 4991 egevaallqeeefdayre-----aln-k-akdyevglvkwdsdydq-----ytleale 5034
 QY 100 VLSQA---METADPLQQL-----VLSAVSGHL-----GKTSDDLKFA----- 135
 DB 5035 wlskteaavgynkldgslqkklvleqfgghlqtlldfwqktdldlmkagvllletcsdt 5094
 QY 136 -----LASPYFVI-----RLEAAYRLANLKNTKYIDHLHSFIRKLPEEI-OC- 176
 DB 5095 risnaqlmqltkynalltlakevmrtembyq-ehqghslsyeecgswiektreklesce 5153
 QY 177 -----LSAATFLRIETESDAYITDILAAKSAIRKATATAOIGTYOQKRRLPT 224
 DB 5154 qipgtlnevgklnltvknllngfgetgnkltlyllelkekvimnteqnnaaklged----- 5208
 QY 225 LRNLTLSPDQDAITAYALGKLMDGOSYVNIKKOLKPDVDVTLAAQAALIAIGK----- 280
 DB 5209 -----tealkgdfkllivdln-----dvyrklian-----rlaqleefklykllle 5249
 QY 281 -EEDALPYIKK-----QALEERPRALYALRHLRPEIG-----IPALPFLFKTKRSEAKLN 330
 DB 5250 wledvepsvksdeflndlsekraalekfrvqrdingndivakngrlkednsldldkd 5309
 QY 331 VALALL---ELGCTPKLLEYITRRLVOPH-----YNETIALSFSKGRQLQNN-KRVNI 380
 DB 5310 fgpqtlkfddqtlqtnkllleel-engynshkkykqayne-----lqgdllrttrri 5357
 QY 381 IVPDDPERERLSTGTGLEQILTFELRPLKEAYLPCITKYLASQKTQATATISFLSH 440

DB 5358 eweq-----cadchekdqvesrlnrlgdigsslegkallaeceelsgavia----- 5405
 QY 441 TSHQBALDLFQAKPLGPEPIIRAYADLAIVMLTKDPPEKRSIADYAKKLIQETL----- 495
 DB 5406 tsqsegqdnvagekl-----hltsewelqglsrdarslslesclawqt 5449
 QY 496 -----LFPVDENQRPHPSPYLRQVTPESRTKMLDLLEFLRLATSKSSQEDRLRLIQ 546
 DB 5450 flqkfnklnlvtetmknrvtksqe--genktpedlvnakkllleevlaekdnvedlnduce 5507
 QY 547 LMTGDA 553
 DB 5508 lmeqsa 5514

RESULT 9
 ABG21017
 ID ABG21017 standard; Protein; 1870 AA.
 AC
 XX ABG21017;
 AC
 XX 18-FEB-2002 (first entry)
 DT
 XX
 DE Novel human diagnostic protein #21008.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PD 30-MAR-2001; 2001WO-US08631.
 PF
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS85204.
 XX
 PT New isolated polynucleotide and encoded polypeptides; useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS
 PS Claim 20; SEQ ID NO 51376; 103pp; English.
 PS
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (I) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1870 AA;

Query Match 4.5%; Score 128.5; DB 22; Length 1870;
Best Local Similarity 19.2%; Pred. No. 0.27;
Matches 102; Conservative 85; Mismatches 154; Indels 189; Gaps 26;

QY 141 PVIRLEAARLANLKNKTYIDHSHFIHLKPE-----IQCLSAIFLRLE--T 187
DB 1341 pvrleaaarlanlknktyidhshfihlkpe-----sflinfihtlengpefsarakvyfaalilvalhglykleyt 1397
QY 188 EESDAYIRDL-----AAKKSAT---RSATVALQIGEQYQOKRFLPTLRNLTLASAPQDEAI 240
DB 1398 dlnhtftellegvyvakpknllrrsvle-----rmlsnwmsi-----c 1439
QY 241 LVALGKLKDGQSYNN---IKKOLQRPVDVTLAAQAALIALGKEEDALPVIRKQALEER 296
DB 1440 lygylkdsagepilyklfkalkhgvkgydav-----q 1472
QY 297 PRALVALRHLRHPSEIGI-----PIALPFIETKNSSEAKLVNALALEGCTPKILEY 348
DB 1473 kkaacyl---ndtgllgdvayapltvsvlyvgdegvd---alpvkvlncdt---lsq 1520
QY 349 ITERLVQPHYN-----ETIALSFSKGRFLQ-----NMKRVNIITVPOD 385
DB 1521 vkekiidqvyrgqscwcpdpdsavlewmpgstaqllsdlclsgregrtwkrvntlmhyn 1580
QY 386 PQRERKLLSTTRGLEQILTLFRLPKEAYLPC-----YKLL-----ASOK 427
DB 1581 vrdgatllskvygvsqgedsqdlpgerhallleenrvwhlvrpctdevdegkskrgsvk 1640
QY 428 TOLATTAISFLSHT-----SHOEALDLLEQAAKLPG-----PLIRAYAD----- 467
DB 1641 ekertkailteilyltrlsvkqtlqgfvdnfigsvlapghavppavxyftifldegaekhn 1700
QY 468 -----LAIR-----NLTKDPEKKRSLH-----DYAKKLIOETLLFVDTENQRP 505
DB 1701 igdedtlhkwktnslprlfrvwnllkphfflfdvhvneavdasvsvalgt---fmdacrite 1758
QY 506 HPSMPLRYQVTPESRRTKMLDLETLANSKSSSD-----IRLLIQLMTEG 551
DB 1759 h-----klrdspsnkllyakeistykkmwedyykgyrrgmvyqsdg 1800

RESULT 10
ABG06301
ID ABG06301 standard; Protein; 2017 AA.
XX
AC ABG06301;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6292.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
PA (HYSR-) HYSRQ INC.
XX
PI Drmanac RT, Liu C, Tang YF;

XX
DR WPI: 2001-639362/73.
DR N-PSTDB; AAS70488.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 36660; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2017 AA;

Query Match 4.5%; Score 127.5; DB 22; Length 2017;
Best Local Similarity 20.0%; Pred. No. 0.36;
Matches 113; Conservative 88; Mismatches 206; Indels 157; Gaps 26;

QY 16 LPISLVAKPPEVSHKTLISTOSTQOALATYLEALDAYDHDHFVLRKIGEDYLKOSIH 75
DB 625 ldvfyvlllqsevalndhlhtkqlkedkvenlvdqlnksqesvnsiqkenle--lkchir 682
QY 76 SSDPQ---TRKSTIIIGGLAGSSSEALDVL---SQAMETADPLQQLVLSAASGHGKTSDD 130
DB 683 qneeeisrltnelmgslngdsnsfkdtllkeraeavrnllkqnseleqinelnkvvaf 742
QY 131 LLEKALASPPVIRLEAARLANLKNKTYIDHSHFIHLKPEEIQCLSAIFLRLETE-- 188
DB 743 v-----xmeneklylacedvnhqlee---clagmqnqlslenntl 778
QY 189 -----ESDAYIRDLAAKSAIRSATALQIGEQYQOK-RFLPTLRNLTLASAPQDEAIL 241
DB 779 vetlkmkgeleaelcwakrrlllea-----nkyekleelnsarnintsaleqlehehl 833
QY 242 VALGKLKDGQSYNNIKKOLQKPDV-----VTLAAQAALIAL-GKEEDALPVIRK-- 289
DB 834 -klnqkxd-metaelkknleqmdcdhketdvjssseegkqtlqllnkketilekixer 891
QY 290 -----KQALEERRPALVALR-----HLPSEIGIPALPFIETKTN 324
DB 892 skrlqeeldkysqalrknellrqtieekdtsjgmkeenhlgee-----lerlr 941
QY 325 SEAKLVNALALEGCTPKLLEYITE-----RLVQPHVETLALSTSKGRFTLQNMK 376
DB 942 eegsrtaeva-----dpkrlidsvtelaseveqnlntikehleeeek-hnqklliedqns 993
QY 377 RVNIITVQDQPERERLLSTTRGLEEQI---LTLFRLPKEAYLPCITYKLLASOKTOIATY 433
DB 994 kmqll--qslqeqkemderyqhegmahthglf-lekdeeklsqktleqiktgl--- 1047
QY 434 AISFLSHTSHOEALD-----LFOAAKLPGEPILIRAYADLAIYNLTKRDEKKRSLDVA 487

```

Db 1048 -----heerqdiqdnsgifgetkvgs-----lniengsek-----hdls 1082
QY 488 K-----KLIQETLLFVDTENOR 504
Db 1083 kaeterlykvikereleikllinek 1106

RESULT 11
AAR99675
ID AAR99675 standard; Protein: 631 AA.
AC AAR99675;
XX
XX 10-OCT-1996 (first entry)
DE RHAMM 1-2a isoform.
XX
XX RHAMM 1-2a; receptor for hyaluronic acid mediated motility;
KW hyaluronan receptor; cell locomotion; cell proliferation;
KW breast cancer; therapy.
XX
XX Mus sp.
OS
XX Key Location/Qualifiers
XX Region 55..79
XX FT /note= "exon 2A-encoded region"
XX PN EP21012-A2.
XX PD 10-JUL-1996.
XX PF 16-OCT-1995; 95EP-0307310.
XX PR 14-OCT-1994; 94GB-0020740.
XX PA (MANI-) MANITOBA CANCER TREATMENT & RES FOUND.
XX PA (UYMA-) UNIT MANITOBA.
XX PI Entwistle J, Turley EA, Zhang S;
XX DR WPI; 1996-310997/32.
XX DR N-PSDB; AAT34525.
XX
XX Receptor for hyaluronic acid-mediated motility protein, and DNA
XX encoding it - useful to treat or prevent diseases associated with
XX the receptor. e.g. breast cancer
XX
XX Claim 8; Page 50-52; 117pp; English.
XX
XX RHAMM 1-2a (AAR99675) is an alternatively spliced variant of
XX RHAMM 1 (AAR99673) (receptor for hyaluronic acid mediated
XX motility), a protein involved in cell locomotion or motility and
XX cell proliferation and transformation. It differs from RHAMM 1
XX by an insertion of 25 amino acids (see also AAR99674) between
XX amino acids 54 and 55 of RHAMM 1, resulting from an alternatively
XX spliced exon 2A (AAT34502). RHAMM 1-2a is the isoform that is
XX overexpressed in tumours. Determination of the level of RHAMM
XX 1-2a in a sample can be used to assess the prognosis of a tumour
XX (esp. breast cancer) patient. The RHAMM 1-2a protein can also
XX be used to suppress or control a tumour by modulating the
XX interaction of cell-associated RHAMM with its ligand.
XX
XX Sequence 631 AA:

```

```

Query Match 4.4%; Score 126; DB 17; Length 631;
Best Local Similarity 21.3%; Pred. No. 0.091;
Matches 115; Conservative 97; Mismatches 213; Indels 114; Gaps 23;

```

```

QY 70 LKOSIHSSDPQTRKSTIR---GAGLAGSSSEALDVLQAMETADPLQQLLVLSAVSGHLGK 126
Db 31 Ikiqatgkdlteskgkivqlegkivsiekehd---ekcetekhlleylqetscsdqvek 87

```

```

QY 127 TSDDLLFKALASPVYIRLEAAVRLANLKNTRVIDHLHSFTIKL--PEEIQCLISAAIFLR 184
Db 88 ckvd-----iaqleedlk---ekdreiliskgleenitfskqiedltvcql- 132
QY 185 LETEESDAYIRD-----LLAAKKSARSTATQIGYX---QOKRFLPTIRNLTLTSRPOD 236
Db 133 lcterdnlvskdraetraetisaemqllterlalerqeyeklqgke-----lgsqsl1q 184
QY 237 QEAILYA--LGRKLDGQSYVYNIKKOLQKPDVDTLAAQALILALGKEEDALPVIRKQALE 294
Db 185 qkeksar1qgqlscfgeemtsekvnfkeekllala---eldavqgkeeserlvkqlee 241
QY 295 ERPRALYALRHLPSLGLPIALPIFLKTRKNSAKLNV-----LALLELGCDTPKLEY 348
Db 242 erkstaqltrldn-----llrekevelekhiahaqa11jaqekyndtaqslrd 291
QY 349 ITERL--VQPHNYETLALSFSGRTLOMMKRVNIIVPODOPRERLSTRTGLEQILTF 406
Db 292 vtaqlsvqekyndt-----agslrd-----vtaqlseqekyndtaqslrd--vla 336
QY 407 LFRLPKEAY-----LPCIVYKLLASOKTQLATATSLFSLSHSNOBALDLFQAARLPGER 460
Db 337 qleseqekyndtaqslrdvtaqlsvqekyndtaqslrdvtaqlses---yksstl---- 388
QY 461 IIRAYADLAIVNLTQDPEKKSLSHDYAKKLIQETLLFVDTENQRPHPSPMYLRVYTPPS 520
Db 389 --kelediklenltl--qekvamaeksvedvqgqlltaestnge-----yarmvqd1qn 438
QY 521 RTKLMLDILEETLATS-----KSSSDIRLLQLMTEGDAKNPVLACGLLIRI 566
Db 439 rstlkeekietstflektidlknglrgqdedftrkgleekykrtaekenvmtelme1 497

RESULT 12
AAR99673
ID AAR99673 standard; Protein: 606 AA.
AC AAR99673;
XX
XX 10-OCT-1996 (first entry)
XX
XX Receptor for hyaluronic acid mediated motility RHAMM 1.
XX
XX RHAMM 1; receptor for hyaluronic acid mediated motility;
XX hyaluronan receptor; cell locomotion; cell proliferation;
XX breast cancer; therapy.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX Modified-site 91..93
XX /label= N-glycosylation_site
XX Modified-site 258..260
XX /label= N-glycosylation_site
XX Modified-site 279..281
XX /label= N-glycosylation_site
XX Modified-site 300..302
XX /label= N-glycosylation_site
XX Modified-site 321..323
XX /label= N-glycosylation_site
XX Modified-site 342..344
XX /label= N-glycosylation_site
XX Modified-site 373..375
XX /label= N-glycosylation_site
XX Modified-site 413..415
XX /label= N-glycosylation_site
XX Modified-site 532..542
XX /label= Hyaluronan_binding_domain-I
XX Modified-site 553..562
XX /label= Hyaluronan_binding_domain-I
XX Modified-site 594..596
XX /label= N-glycosylation_site

```



```

QY 70 LKOSIHSSDPOTRSTII---GAGLAGSSEALDVLSQAMETADPLQOLLVLSANGHICK 126
Db 31 lkgitgkdileskgyqlgkivlsiekid---ekcefeklleylqelscadqyek 87
QY 127 TSDLLFKALASPPYVIRLEAAYRLANKTKVFDHLSFIKFL--PEEIOCLSAIFLR 184
Db 88 ckvd-----laqleedlk---ekdreilslksleentlfsqldeltvkq- 132
QY 185 LETESDAVIR-----LLAAKSAIRSATALQIGEY---QOKRPLPTLRNLITSASPOD 236
Db 133 lcterdnlvskdreraetlisaemqllterlalerqeyklqke-----lgsqslq 184
QY 237 QEALIIYA--LGKIKDGSQYNNIKQLOKPDVDVTLAAQALIALGKEDALPVIKKOLE 294
Db 185 gkeksaralqgqlcfsgeemsekvnfkeelkiala---eldavqkseeqserlvkqlae 241
QY 295 ERPRLALYRLHPSEIGIPALPIFLKTKNSEAKLNVA-----LALLELGCDPFLLEY 348
Db 242 erkstaegltirld-----llrekevelekhiahaqalljaekyndtagslrd 291
QY 349 ITERL--VQPHYNETLALSFSGRGLQMKRNVIIVPDQPERKLLSTTKGLEQILTF 406
Db 292 vtaqlasvgekyndt-----agslrd-----vtaqlaseqekyndtagslrd--vta 336
QY 407 LFRLPKEAY-----LPCITVKLASQKTQLATTAISFLSHTSQHALDLLFOAKLPGE 460
Db 337 qleseqekyndtagslrdvtaqlaseq-kyndtagslrdvtaqlaseq---yasscl----- 387
QY 461 IIRAVADLAIVNLTKDPKKRSLHDYAKKLIQETLLFVDTEQRPHPMPYLKRVGPES 520
Db 388 --kelediklenltl--qekvamaeksvedvqgqlltaesnge-----yarmvqdlqn 437
QY 521 RTKMLDLLETLANS-----KSESDIRLLIQLMTEGDAKFPVLAGLLIKI 566
Db 438 rstlkeekietssflektidlnqldgdedftfkgleekykrtaekenmtlmei 496

RESULT 14
AAU00015
ID AAU00015 standard; Protein: 1838 AA.
XX
AC AAU00015;
XX
DT 09-MAY-2001 (first entry)
XX
DE Human Plexin-B2.
XX
KW Human; Plexin-B2; semaphorin domain; hyperplasia; neoplasia; cancer;
KW neurodegenerative disease; autoimmune disease; lupus; multiple sclerosis;
KW inflammatory bowel disease; diabetes type I; rheumatoid arthritis;
KW immunogen; antibody; MET related sequence; MRS.
XX
DS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT FT /label= Signal-peptide
FT FT 1..1199
FT FT /label= "Extracellular domain"
FT FT 19..1838
FT FT /label= "Mature Plexin-B2"
FT FT 19..518
FT FT /label= "Semaphorin domain"
FT FT 451..530
FT FT /label= "Primary Met related sequence, MRS"
FT FT 601..680
FT FT /label= "Secondary MET related sequence, MRS"
FT FT 751..830
FT FT /label= "Tertiary MET related sequence, MRS"
FT FT 800..1010
FT FT /label= "G-P repeat region"
FT FT 1196..1215
FT FT /label= "Transmembrane domain"

```

```

XX XX WO200114420-A2.
FN FN
XX XX
PD PD
XX XX 01-MAR-2001.
XX XX
PF PF 25-AUG-2000; 2000WO-US23365.
XX XX
PR PR 25-AUG-1999; 99US-0150576.
XX XX
PA (UYTO-) UNIV TORINO.
PA (REGC) UNIV CALIFORNIA.
XX XX
PI Artigiani S, Comoglio PM, Goodman CS, Tesier-Lavigne M;
PI Tamagnone L;
DR WPI: 2001-226610/23.
DR N-PSDB: AAS00020.
XX XX
PT New plexin polynucleotides and polypeptides, useful in diagnosis,
PT therapy and in producing compounds for treating diseases involving
PT aberrant cell growth (e.g. cancer) or immune regulation (e.g.
PT autoimmune diseases)
PS PS
PS Claim 4: Page 52-56; 79pp; English.
XX XX
CC The sequence represents Human Plexin-B2. Plexins are large transmembrane
CC proteins whose extracellular domain shares homology with Scatter factor
CC receptors and contain an approximately 500 amino acid Semaphorin domain.
CC The plexin polynucleotides and polypeptides, and plexin-specific binding
CC agents are useful in diagnosis, therapy and in the biopharmaceutical
CC industry. In particular, the plexin polynucleotides and polypeptides are
CC useful for generating compounds (e.g. plexin-specific binding agents or
CC antibodies) for treating or diagnosing a disease or disorder involving
CC aberrant cell growth (e.g. hyperplasia, neoplasia, cancer or
CC neurodegenerative disease), or diseases or disorders involving aberrant
CC immune regulation (e.g. autoimmune diseases such as lupus, inflammatory
CC bowel disease or diabetes Type I), or immunosuppressive diseases such as
CC multiple sclerosis or rheumatoid arthritis.
XX XX
S0 Sequence 1838 AA;

Query Match 4.3%; Score 121.5; DB 22; Length 1838;
Best Local Similarity 18.9%; Pred. No. 1;
Matches 98; Conservative 84; Mismatches 162; Indels 175; Gaps 23;

QY 141 PVIRLEAAYRLANKTKVFDHLSFIKFLPEE-----IQCSAIFLRL--T 187
Db 1310 pve-galyqgsnlnsk--sflnflhtlenqreifsarakvyaslltvalhgkleyt 1366
QY 188 EESDAVIRDLAANKSAIRSATALQIGEQOKRPLPTLRNLITSASPODEAIIYALGKL 247
Db 1367 dimhtlfielleqyvvavknplmlirsetverlmsnmsl-----clgykld 1415
QY 248 KDGSQSYN-----IKQLOKPDVDVTLAAQALIALGKEDALPVYKQALEERRALYAL 303
Db 1416 sageplyklfkaklkhyekgvdav-----qkkakycl 1448
QY 304 RHLPSEIGI-----PIALPIFLKTKNSEAKLNVALALLEGDDPKLLEYITERVQ 355
Db 1449 -----ndgllgddveyapilvsvivgdegvd-----alpvvlnlndt--lsqyekltd 1496
QY 356 PHYN-----ETLALSFSGRGLQ-----NMKRVNIIVPDQPERRL 392
Db 1497 qyrgqpcscwprpdsvvlewprpstqllsdldltsqregvrvrvntlmynvrtdgatl 1556
QY 393 LSTRGLEEQILTFRLPKAVYLP-----TYKLL-----ASQKQLATTA 434
Db 1557 llskvgyvsgqpedsqdqlperhalileenvvhlvprtdvdegkkskrgsvkekerka 1616
QY 435 ISFLSHT-----SHQEALDLLFOAKLPGE-----PIIRAVAD-----L 468
Db 1617 lteilyltrl]lsvyglqlqgvdnftfsvlapyhavpavkykffdlfidegaekhnlqdedti 1676

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 03:16:15 ; Search time 24.01 Seconds
(without alignments)
577.832 Million cell updates/sec

Title: US-09-662-812-2
Perfect score: 2833
Sequence: 1 MGLPHTLFGILLCSLPISL.....TEGDANKPPVLAGLLIKIVE 568

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121.5	4.3	2987	2	US-08-970-269A-29
2	121.5	4.3	2987	4	US-09-407-562-29
3	121.5	4.3	3959	2	US-08-970-269A-30
4	121.5	4.3	3959	4	US-09-407-562-30
5	114	4.0	1388	2	US-08-685-576-4
6	113	4.0	2101	1	US-08-195-487-4
7	113	4.0	2101	5	PCT-US93-06160-4
8	112	4.0	860	1	US-08-117-362-3
9	112	4.0	860	1	US-08-486-924-3
10	112	4.0	1312	2	US-08-592-126-148
11	112	4.0	1312	2	US-08-687-080-51
12	111	3.9	2101	1	US-08-466-390-4
13	111	3.9	2101	1	US-08-470-950-4
14	111	3.9	2101	1	US-08-467-781-4
15	111	3.9	2101	2	US-08-483-924-4
16	111	3.9	2101	4	US-09-452-294-1
17	110.5	3.9	442	4	US-09-081-689-2
18	109	3.8	582	1	US-08-261-086-2
19	109	3.8	582	1	US-08-261-086-4
20	109	3.8	3056	1	US-08-508-836A-8
21	109	3.8	3056	2	US-08-629-001A-3
22	109	3.8	3056	2	US-08-874-266-2
23	109	3.8	3056	4	US-08-642-274D-3
24	109	3.8	3056	4	US-08-642-274D-3
25	109	3.8	3056	4	US-08-952-127-3
26	108	3.8	1332	2	US-08-952-014C-3
27	108	3.8	1332	2	US-08-971-244-2
					Sequence 2, Appl

28	106	3.7	582	1	US-08-261-086-6	Sequence 6, Appl
29	106	3.7	1257	2	US-08-750-152A-2	Sequence 2, Appl
30	105	3.7	477	1	US-08-402-217A-3	Sequence 3, Appl
31	105	3.7	477	1	US-08-700-178-3	Sequence 3, Appl
32	105	3.7	477	1	US-08-995-654-3	Sequence 3, Appl
33	105	3.7	1866	4	US-08-938-105-3	Sequence 3, Appl
34	104.5	3.7	2629	2	US-08-751-189-4	Sequence 4, Appl
35	104.5	3.7	2629	2	US-09-060-836-4	Sequence 4, Appl
36	104.5	3.7	2629	2	US-09-184-445-4	Sequence 4, Appl
37	104	3.7	504	3	US-09-231-529-6	Sequence 6, Appl
38	104	3.7	504	4	US-08-977-816-6	Sequence 6, Appl
39	103.5	3.7	976	4	US-09-104-324B-4	Sequence 4, Appl
40	103	3.6	894	4	US-08-117-362-4	Sequence 4, Appl
41	103	3.6	894	4	US-08-486-924-4	Sequence 4, Appl
42	102.5	3.6	835	2	US-08-968-751-4	Sequence 4, Appl
43	101.5	3.6	744	1	US-08-764-100-25	Sequence 25, Appl
44	101.5	3.6	2482	1	US-08-328-254-6	Sequence 6, Appl
45	101.5	3.6	3248	1	US-08-353-700-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-970-269A-29
; Sequence 29, Application US/08970269A
; Patent No. 5976803
; GENERAL INFORMATION:
; APPLICANT: Katheryn Meek
; TITLE OF INVENTION: Genetic Test For Equine Severe
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970,269A
; FILING DATE: No. 5976803ember 14, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler Ph.D., Benjamin A.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5860
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2987 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; FEATURE:
; US-08-970-269A-29

Query Match 4.3%; Score 121.5; DB 2; Length 2987;
Best Local Similarity 21.4%; Pred. No. 0.16;
Matches 128; Conservative 83; Mismatches 196; Indels 191; Gaps 32;


```

; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970,269A
; FILING DATE: No. 5976803ember 14, 1997
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler Ph.D., Benjamin A.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5860
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908
; INFORMATION FOR SEQ ID NO: 30:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3959 amino acid
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: no
; ANTI-SENSE: no
;
; US-08-970-269A-30

```

Query Match 4.3%; Score 121.5; DB 2; Length 3959;

Best Local Similarity 21.4%; Pred. No. 0.26;

Matches 128; Conservative 83; Mismatches 196; Indels 191; Gaps 32;

```

QY 31 KILYISTQST-QOALATYVLEALDYGHDFFVLKRIKIGEDYLGKSIHSSDPQTR----- 82
DB 1107 KVLGTERQSSLMKAVAFLESI---AMHDIMA---AKYFGTATGNRPSPQGERYNY 1159
QY 83 -KSTIIAGLAGSSEALDVLVSQAMETADPLQOLLVLSAVSGHLKTSDDLFKALASFPY 141
DB 1160 SKCTIV-----VRIMEFTTLLSTSPGKLLKEDVCNTNMK-----LLVKTLCPESS 1208
QY 142 VIRLEAAYRLANKNTYVIDLHSFIKLPREICLSAIFLRLETESDAYIRDLAAK 201
DB 1209 I-----GENIDVAVMNLPSVCNTL-----MKALK 1234
QY 202 KSAIRSATVALQIGEYQOKRFLPTLRNLTSASPODOEAI-LYALGKLKDG---QSYNYNIK 257
DB 1235 KSPYKDIEMH-----LKEKITAQSIELCAVDLYCPDACVDRARLASVVSAC 1282
QY 258 KQLOKRPVDVTLAAQAL-----IALGKEEDALPYI---KKQA--LE 294
DB 1283 KQLRACAVLCITIPSGSDQDHSIGTKLLSLVYKSIAPGSDQQLPLSDIPCKRLASGLL 1342
QY 295 ERPALALAL-RHLPSEI-GIPIALP-----IFLTKTNSAKINLV 331
DB 1343 ELAFAPGLCEHLVSLDLDTTVLSPSRGSGOKNIVSFSHGEPYFSFSEIETINTELLKNL 1402
QY 332 ALALLEI---GCOFPKLLLEY-----TERLVOPHYNNTLALSCKGKITONMKRVN 379
DB 1403 DLAVLELMKSSVDPKPKVSNVNGMLDQSFDRRTSEKHQGLKATIT-----LQNMKKCD 1457
QY 380 IIVPQDERRLITSTRGLEEOILTFPLPKAEAYL---PCIYKLASOKTQATTAIS 436
DB 1458 SMMAKDSAPESKMAVL-----LAKIFQIDSSVCFTNHCW-----PEVFTIYVS 1504
QY 437 FLSTHTSQEALDLFOAKALPGEPIIRAYADLAIV-NLTDPEKKRSIHDAKLIQETL 495
DB 1505 LLAIDSK---LDL-----HLKGQAILI-----LPFTSTLTGG-----SLEDL--KVVDENL 1544

```

```

QY 496 L---FVDFENORPHPSMYLRYQYTPESRTKLMIDILETLATSKSSEDIIRLLQMT 550
DB 1545 IVSNFPMKSEEPPTQYNNY-----VDMKKFIDALELSKSP-----MLDQMT 1591

```

RESULT 4

US-09-407-562-30

Sequence 30; Application US/09407562

Patent No. 6294334

GENERAL INFORMATION:

APPLICANT: Kathryn Week

TITLE OF INVENTION: Genetic Test For Equine Severe

TITLE OF INVENTION: Combined Immunodeficiency Disease

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dr. Benjamin A. Adler

STREET: 8011 Candle Lane

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple

OPERATING SYSTEM: Macintosh

SOFTWARE: Microsoft Word for Macintosh

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/407,562

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/970,269

FILING DATE: No. 6294334ember 14, 1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Adler Ph.D., Benjamin A.

REGISTRATION NUMBER: 35,423

REFERENCE/DOCKET NUMBER: D5860

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-777-2321

TELEFAX: 713-777-6908

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 3959 amino acid

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: no

ANTI-SENSE: no

US-09-407-562-30

Query Match

4.3%; Score 121.5; DB 4; Length 3959;

Best Local Similarity 21.4%; Pred. No. 0.26;

Matches 128; Conservative 83; Mismatches 196; Indels 191; Gaps 32;

```

QY 31 KILYISTQST-QOALATYVLEALDYGHDFFVLKRIKIGEDYLGKSIHSSDPQTR----- 82
DB 1107 KVLGTERQSSLMKAVAFLESI---AMHDIMA---AKYFGTATGNRPSPQGERYNY 1159
QY 83 -KSTIIAGLAGSSEALDVLVSQAMETADPLQOLLVLSAVSGHLKTSDDLFKALASFPY 141
DB 1160 SKCTIV-----VRIMEFTTLLSTSPGKLLKEDVCNTNMK-----LLVKTLCPESS 1208
QY 142 VIRLEAAYRLANKNTYVIDLHSFIKLPREICLSAIFLRLETESDAYIRDLAAK 201
DB 1209 I-----GENIDVAVMNLPSVCNTL-----MKALK 1234
QY 202 KSAIRSATVALQIGEYQOKRFLPTLRNLTSASPODOEAI-LYALGKLKDG---QSYNYNIK 257
DB 1235 KSPYKDIEMH-----LKEKITAQSIELCAVDLYCPDACVDRARLASVVSAC 1282

```

QY 258 KOLKRPDVTYLAQAAL-----IALGKEEDALPVI---KKA--LE 294
| | | : : : : : | | | : : : : :
Db 1283 KOLHRAVGLCVIIPQADQHHISGKTLSTLYKSIAPDEQOCLPSLDPNCKRLASGLL 1342
| | | : : : : : | | | : : : : :
QY 295 EPRRLVYL-RHLPEI--GIPALP-----IFLKTNSAKLVN 331
| | | : : : : : | | | : : : : :
Db 1343 ELAFAPGGLCEHLVSLDITVLSMPSRSGSQKNIVSFSGHYEYSLFSETINTELKML 1402
| | | : : : : : | | | : : : : :
QY 332 ALALHEL---GCDTFLLEYI-----TERLVAPHYNETLASFSGKRTLOMKRVN 379
| | | : : : : : | | | : : : : :
Db 1403 DLAVLELMKSSVDNPKMNSVNLGMLDOSFRDRTSEKHGKLLATII-----LQWKKCD 1457
| | | : : : : : | | | : : : : :
QY 380 IIVPODPOERERLSTTGLEQIILFLFRPEAYL---PCIVKLASQKTQATTAIS 436
| | | : : : : : | | | : : : : :
Db 1458 SMMWADSAPESKMAVLT-----LLAKIFQIDSSVCFTNHCMF-----PEVFTYVS 1504
| | | : : : : : | | | : : : : :
QY 437 FLSSHQEAIDLFLQAAKLPGEPIIRAVADAIY-NILTKDEKRSLLHDYAKKLIQETL 495
| | | : : : : : | | | : : : : :
Db 1505 LLAASK---LDL-----HLKGAAILL---LPFTSLTG---SLEDL--KVVLEML 1544
| | | : : : : : | | | : : : : :
QY 496 L---FVDTENORHPSPMYLRYQVTPESRTKMLDILETLATSKSSEDIRLIQLMTE 550
| | | : : : : : | | | : : : : :
Db 1545 IYSNPMKSEEPPTGLQNNY-----VDCMKKFLDALSLKSP---MLQLMTE 1591
| | | : : : : : | | | : : : : :

RESULT 5
US-08-685-576-4
; Sequence 4, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Koza
; APPLICANT: Iwamoto, Akihito
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-4

Query Match 4.0%; Score 114; DB 2; Length 1388;
Best Local Similarity 20.2%; Pred. No. 0.23;
Matches 134; Conservative 91; Mismatches 258; Indels 180; Gaps 27;

QY 52 DAYGDHDFVLRK-----IGEDYLKQSTHSSD-PQTRKSTIIAGLASSSEAL--- 98
| | | : : : : : | | | : : : : :
Db 399 DKGDFETFPKPAVGNQLPEIFGTYRENLILSDSPSCRNDISQSRKNEESQIOKK 448
| | | : : : : : | | | : : : : :
QY 99 -----DVLSQAMETADPLQQLLVLSAVSGHLKTSDDL-----LPKALAPPVY-RLEA 147
| | | : : : : : | | | : : : : :
Db 449 LYTEEHLNSNQAKLEELQ--KCKSVNRLKRTAKELEFEETLLKSVSALROLEREA 506
| | | : : : : : | | | : : : : :
QY 148 AYRLANLKNKYVDH-----LHSFIKLPKEIQL-----SAALFLRETEE 189
| | | : : : : : | | | : : : : :
Db 507 LLQHKNAEYQKADHEADKRLNDVNSLKDQLEDLKRQNSQISTEKVNLQRLQDE 566
| | | : : : : : | | | : : : : :
QY 190 SDAYIR--DLAAKSATSATLQIGEQKRPFLTNLNLTSASPDQEAIIYALGL 247
| | | : : : : : | | | : : : : :
Db 567 TNALRTSDTRARLRKTKQAESKQIQLESNN-----RDL-----ODKNCL-ETAKL 614
| | | : : : : : | | | : : : : :
QY 248 KDGQSYNNIKQOLQKRPDVT-----LAAQALALGKEED-----ALPVRKQALE 294
| | | : : : : : | | | : : : : :
Db 615 KLEKEFINQSLSEBRDRTHGSELINDQGRIC-GLEEDLKNQKILLAKVELEKRLQ 673
| | | : : : : : | | | : : : : :
QY 295 EPRRL-----YALR-----HLPEIGIPALPFLKYN-----SEAK 328
| | | : : : : : | | | : : : : :
Db 674 EFTDLEREKSNMEIDMYOLKVIQOSLEOEAEHKATKARLADNKKIYESIEAKSEAM 733
| | | : : : : : | | | : : : : :
QY 329 LNVALLALE-----IGCDPFLLEYITTEBYQPHY----- 358
| | | : : : : : | | | : : : : :
Db 734 KEMEKLLLEERTLKQKVENLLLEAKRCSLLDCDQKQSOQKINELIKQDVLNEDVRLT 793
| | | : : : : : | | | : : : : :
QY 359 ----NETLALSFSGKRTLOMKRVNII---VQDPOERERLSTTGLEE----- 401
| | | : : : : : | | | : : : : :
Db 794 LKIEQETQKRCILQNDLKNQOVNTLKNSEKQLQENNHLMEMKNLEKQVLAELKERO 853
| | | : : : : : | | | : : : : :
QY 402 ----QILFLFLPKREAVLPCTYKLL-----ASQKTOLATTAISFHSHTOE----- 445
| | | : : : : : | | | : : : : :
Db 854 DADGOMKELODLEAEQVFTLYKTVREKCECEKTKLGEKEL-----QKKKQELQDER 908
| | | : : : : : | | | : : : : :
QY 446 ---ALDLLFQAAKLPGEPIRA-----YADLAIYMLTPOPEKRSLLHDYAKKLIQETLLE 497
| | | : : : : : | | | : : : : :
Db 909 DSLAQOLEITTLTKADSEOLANSIAEQVSDLEKEKIMKELEIKEMARHKQELTEKDATI 968
| | | : : : : : | | | : : : : :
QY 498 VDTENORHPSPMYLRYQVTPESRTKMLDILETLATSKSSEDIRLI-----QLMTEG 551
| | | : : : : : | | | : : : : :
Db 969 ASLEETNRLTSDVANLANNEKELNNKLDVQEQSLKDEITSAALKAQERQELTER 1028
| | | : : : : : | | | : : : : :
QY 552 DAK 554
| | | : : : : : | | | : : : : :
Db 1029 TLK 1031
| | | : : : : : | | | : : : : :

RESULT 6
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOKURATY, GARY
; APPLICANT: LINDARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HUMWITZ & THIBEAULT
; STREET: 53 STATE STREET


```

: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/195,487
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/901,701
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MTP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/248-7000
: TELEFAX: 617/248-7100
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-195-487-4

```

Query Match 4.0%; Score 113; DB 1; Length 2101;

Best Local Similarity 21.9%; Pred. No. 0.56; Matches 113; Conservative 81; Mismatches 184; Indels 138; Gaps 25;

```

QY 38 QSTQOALATYIEALDAYDHDHFVLRKIGEDY-LKQSIHSSDPQTRKSTIGAGIAGSSE 96
D 1387 QSKQAGGLRAELRLAQ-----RELGEIIPLRQKVAED---RTAQQLRAEKASYAE 1435
QY 97 ALDVLQAMETADPLQQLLVLSAVSGHLGKTS-----DLFKALASPPVIRLE 146
D 1436 QLSMLKKAHG-----LLAEENRGLGERANLGRQFLEVLELDQAREKYYQELAAVRAD 1486
QY 147 AAYRLANLK-----NTRVIDHLSFTHKLPBEETQCSAAIFLRLE 186
D 1487 AETRLAEVQRAQSTARELEVMTAKYEGAKYKVLLEERQF---QEERQKILTA-----Q 1536
QY 187 TEESDAYIRDLAALKKASIRSATAIQI--GEYQO--KRFPLTRLRLTILTSASPODEAILY 242
D 1537 VEELSKRLADSDQASKYQOQKLKAVQAOGSGSQDEAQRFOQLNELQAOLSKQEDAAEHY 1596
QY 243 ALGKIKDQSYVNIKK-----OLQKPDVDTLAAQAALITALGKEEDALPVI 288
D 1597 KL-QMEKAKTHYDAKKQOQNOELOERLRLSELQLOKENEKELRAEAER---LGHELDQAQGLK 1651
QY 289 KKQALEERPRLALYA-LRHLPSSEIG-----IPALPLFLTKNSEAK--LNVALA 334
D 1652 TKEA-EQTCRHILTAQVRSLEAQVAHADQQLRDLGKFQVATDA-LKSREPQAKPOLDLISID 1709
QY 335 LLELGCDPRLLEVTTELVQPHVNETIALSFSKGRITLQNMKRVNIIVPOPOE--RRRL 392
D 1710 SLDSLCEGTPLS-ITSKLPTQPDGT-----SYGPEPASISQRL 1749
QY 393 LSTRGLEGQILTFV---FRLPKAEVLPICIVKLL--ASQKTQALATTAISFHTSHQAL 447
D 1750 PPKEYESTLEVTPLIPASQAPLESLSIDSLGVFLDSGRKTRSAARRTQTIIINIMTKKL 1809
QY 448 DLFLQAAKLPGEPITIRADIAIYNLTDPPEKKRSL 483
D 1810 DV-----EPP--DSANSSFSYSTRASAPASQSL 1834

```

RESULT 7

PCT-US93-06160-4

; Sequence 4, Application PC/TUS9306160

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE

; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TESTA HUMWITZ & THIBEAULT

; STREET: 53 STATE STREET

; CITY: BOSTON

; STATE: MA

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/06160

; FILING DATE: 19930621

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: PITCHER ESQ, EDMUND R

; REGISTRATION NUMBER: 27,829

; REFERENCE/DOCKET NUMBER: MTP-013

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/248-7000

; TELEFAX: 617/248-7100

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2101 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US93-06160-4

Query Match 4.0%; Score 113; DB 5; Length 2101;

Best Local Similarity 21.9%; Pred. No. 0.56; Matches 113; Conservative 81; Mismatches 184; Indels 138; Gaps 25;

```

QY 38 QSTQOALATYIEALDAYDHDHFVLRKIGEDY-LKQSIHSSDPQTRKSTIGAGIAGSSE 96
D 1387 QSKQAGGLRAELRLAQ-----RELGEIIPLRQKVAED---RTAQQLRAEKASYAE 1435
QY 97 ALDVLQAMETADPLQQLLVLSAVSGHLGKTS-----DLFKALASPPVIRLE 146
D 1436 QLSMLKKAHG-----LLAEENRGLGERANLGRQFLEVLELDQAREKYYQELAAVRAD 1486
QY 147 AAYRLANLK-----NTRVIDHLSFTHKLPBEETQCSAAIFLRLE 186
D 1487 AETRLAEVQRAQSTARELEVMTAKYEGAKYKVLLEERQF---QEERQKILTA-----Q 1536
QY 187 TEESDAYIRDLAALKKASIRSATAIQI--GEYQO--KRFPLTRLRLTILTSASPODEAILY 242
D 1537 VEELSKRLADSDQASKYQOQKLKAVQAOGSGSQDEAQRFOQLNELQAOLSKQEDAAEHY 1596
QY 243 ALGKIKDQSYVNIKK-----OLQKPDVDTLAAQAALITALGKEEDALPVI 288
D 1597 KL-QMEKAKTHYDAKKQOQNOELOERLRLSELQLOKENEKELRAEAER---LGHELDQAQGLK 1651
QY 289 KKQALEERPRLALYA-LRHLPSSEIG-----IPALPLFLTKNSEAK--LNVALA 334
D 1652 TKEA-EQTCRHILTAQVRSLEAQVAHADQQLRDLGKFQVATDA-LKSREPQAKPOLDLISID 1709
QY 335 LLELGCDPRLLEVTTELVQPHVNETIALSFSKGRITLQNMKRVNIIVPOPOE--RRRL 392
D 1710 SLDSLCEGTPLS-ITSKLPTQPDGT-----SYGPEPASISQRL 1749
QY 393 LSTRGLEGQILTFV---FRLPKAEVLPICIVKLL--ASQKTQALATTAISFHTSHQAL 447
D 1750 PPKEYESTLEVTPLIPASQAPLESLSIDSLGVFLDSGRKTRSAARRTQTIIINIMTKKL 1809
QY 448 DLFLQAAKLPGEPITIRADIAIYNLTDPPEKKRSL 483
D 1810 DV-----EPP--DSANSSFSYSTRASAPASQSL 1834

```

Db 1750 PKVESLESLYFTPIPARBQALPESLDSIGVFLDSGKRTSARRRTTQIINITTKKL 1809
Qy 448 DLLEQAKLPGEPIIRAYADLAIYNLTCKDPEKKRSL 483
Db 1810 DV-----EEP---DSANSPFYSTRSAPASQASL 1834

RESULT 8
US-08-117-362-3
; Sequence 3, Application US/08117362
; Patent No. 5595872
; GENERAL INFORMATION:
; APPLICANT: Wettlau II, John R.
; APPLICANT: Sharp, Daru Y.
; APPLICANT: Gregg, Richard E.
; TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08-117,362
; FILING DATE: 03-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC21b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-4526
; TELEFAX: (609) 252-5901
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 860 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-117-362-3

Query Match 4.0%; Score 112; DB 1; Length 860;
Best Local Similarity 19.8%; Pred. No. 0.16;
Matches 127; Conservative 98; Mismatches 226; Indels 192; Gaps 32;

Qy 30 HKILYISTOSTQOALATYIEALDAGDHDFYLRKIGEY-----LKSHSSDPQTK 83
Db 171 HOVLGVTSKAT-----SVTTYKIDSEFYVAVLSSEIRALRNLFQSI----- 212

Qy 84 STIIAGLAGSSEALDV-LSQAMETADPLQOOLVLSAVSGHLGKTSDDLFEALASPPV 142
Db 213 -----AGKIVSRQKLEKTEESVRLKRGKVAA-----ITKAVDSKYTA 252

Qy 143 IRLAAAYRLANKN-TKYVIDLHSP-IRKLPEIQLSAA-IFLRLEESDAYIRDLA 199
Db 253 IPIVGQVFSQKCGPSEHWSIRKHLQPNLSKAEAVRSFL-----AFIKHLRT 304

Qy 200 AKKSAIRSATALQIGEQKRLPTLRNLISASPOD-----QEAR 240
Db 305 AKKEEI-----QIILKAEKEVLPOLVDVAVTSAGTPSDILDLDFEKSTESVILQERF 359

Qy 241 LYA-----LGRKLDGOSYVNIKKOLKOPVDVDTLAA----- 271

Db 360 LYACAFASHPDEELRALLSKRFSGSNDIRE-----SVMIIGALVYRKLCONOGCKIK 414
Qy 272 ----AQAALALGKEDALPVIRKQALEERPRALYARHLPEIGIPALPIFLK-TKNSE 326
Db 415 GVEAKKLLIGLGE-----KAEKKEDIVMYLLAKN-----ARLEGIPLILKYTEGE 463

Qy 327 AKLNVALLLELGGCTPKLLEYITERLVQ-----HYNETL-----ALSFSGKRT 371
Db 464 GPISHLAATTLQRYDP-----FTTDEVKRTMNRIFYHQNKRIHEKTVRTAAALITLKNPS 519

Qy 372 LQMKRVNIIIVPODERELLSTTRGLEOILTFRLRPKEAYLPCIVKLAS----- 425
Db 520 YNEVKNILLISIGELPKEMKMYLS---IYQDILR--FETPASMVQVLKEMVAHNYDF 574

Qy 426 QKTQATTAISPLSHTSHOE--ALDLFQAAKLPGEPIIRAYADLAIYNLTCKDPEKKRS 482
Db 575 SKGSSSAVGYERTSHASYSIDLVS-----GSGILRR-SNLNIFYI-----EKTP 624

Qy 483 LHDYAKKLLQEL--LFVYT-----ENQRPHPSPYLRY--QYTPESRTKMLDILETLA 533
Db 625 LHGIQVIEAOGLEALIAATPDEGEENLDSYGLSALLPFDVQLRPVTFENGYSDLMSKML 684

Qy 534 TSKSS-----EDIRLLQLMTE-----GDAKNPPVLAGLLIKI 566
Db 685 SASDPMGVYKGLLLIDHSQELQDSGLKANMDVGGGLAIDI 727

RESULT 9
US-08-486-924-3
; Sequence 3, Application US/08486924
; Patent No. 5789197
; GENERAL INFORMATION:
; APPLICANT: Wettlau II, John R.
; APPLICANT: Sharp, Daru Y.
; APPLICANT: Gregg, Richard E.
; TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08-486,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/117,362
; FILING DATE: 03-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC21b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-4526
; TELEFAX: (609) 252-5901
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 860 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-486-924-3

Query Match 4.0%; Score 112; DB 1; Length 860;
 Best Local Similarity 19.8%; Pred. No. 0.16;
 Matches 127; Conservative 98; Mismatches 226; Indels 192; Gaps 32;

QY 30 HKLIYSTOSTQOALATYVLEALDAYGHDHFFVLKIGEDY-----LKOSHSSDPQTRK 83
 Db 171 HOVLGVYSKAT-----SVTTYKIEDSFVAVLSEETRALRNLNFOST----- 212
 QY 84 STIAGLASSSEALDY-LSOAMETADPLOQLLVLSAVSGHGTSDDLLEKALASPPV 142
 Db 213 -----AKIVSRQKLEKTTTEASVRLKPGKOVAA-----ITKAVDSKYTA 252
 QY 143 IRLEAAVRLANLKN-TKVIDHLHSF-IHKLPREEIOCLSA-IFPLRETESDAYRDLA 199
 Db 253 IPIVGVFGQSKKCCPSLSHMQSIRKHLDPDNLSKAEAVRSFL-----ATIKHLRT 304
 QY 200 AKSAISRATALQIGEQYQKRFPLETLNLLTSASPOD-----OEAI 240
 Db 305 AKKEEI-----LQILKAENKEVLPQVLDAVTSQTPSDLAIDLEDFEKSTESVILQERF 359
 QY 241 LYA-----LGKLDGQSYNNKKQLQKPDVDVTLA----- 271
 Db 360 LYACAPASHDEELLRALISKFGSGSNDIRE-----SWITIGALVRKICNOGCKLK 414
 QY 272 -----AQAALLAGKEEDALPYIKQALEERPRALYALRHPSELGIPIALPIFLK-TKNSE 326
 Db 415 GVIEAKKLIIGGLE-----KAKKEDIVWYLLALKN---ARLPGLIPLLAKTYTEGE 463
 QY 327 AKLNVALALLEGCDTPKLEYITERLVOP-----HYNEFL-----ALSFSKGR 371
 Db 464 GPISHLAATLQRYDVP-----FITDEVKKTMMNRIYHONRKTHEKTVRTAAAILKNPS 519
 QY 372 LQNNKRNITVPODPOERERLLSTTRGLEQILFELRPLEAVLPQCYKLAS----- 425
 Db 520 YMEKKNLLISGELPKEMKNYMLS---IVQDILR--PETPASKVAVROYLKMAVHNDRF 574
 QY 426 OKTQALATTAISFLSHTSHQE---ALDLLFOAKLPGEPPIIRAVADLAIVNLTCKPEKRS 482
 Db 575 SKSGSSAVTYGYERTSHSASTYSLDILYS-----GSGILRR-SNLNIFQYI---EKP 624
 QY 483 LHDYAKKLIQETL--LVVDI-----ENORPHPSKPYRLY--QVTPESRTKIMLDILETLA 533
 Db 625 LHGIQVIEAQGLEALTAAPPEDEGENIDSYAGLSALLFVQLRPVIFENGSDLSKML 684
 QY 534 TSKSS-----EDIRLLIQLMTE-----GDAKNFPVLAGLLIKI 566
 Db 685 SASDPMYSVYKGLLLIDHSOELQLOSGLXANMDVGGGLADI 727
 RESULT 10
 US-08-592-126-148
 ; Sequence 148, Application US/08592126
 ; Patent No. 5821091
 ; GENERAL INFORMATION:
 ; APPLICANT: Gregory Dolganov
 ; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
 ; NUMBER OF SEQUENCES: 151
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dehlinger & Associates
 ; STREET: 350 Cambridge Avenue, Suite 250
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/592.126
 ; FILING DATE:

CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Sholtz, Charles K.
 REGISTRATION NUMBER: 38,615
 REFERENCE/DOCKET NUMBER: 4600-0111
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 324-0880
 TELEFAX: (415) 324-0960
 INFORMATION FOR SEQ ID NO: 148:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1312 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 INDIVIDUAL ISOLATE: Rad50, pro-translation of SEQ ID NO:54
 US-08-592-126-148

Query Match 4.0%; Score 112; DB 2; Length 1312;
 Best Local Similarity 20.6%; Pred. No. 0.32; Indels 132; Gaps 25;
 Matches 114; Conservative 91; Mismatches 216;

QY 62 LRKIGEDYLKOSHSSDPQTRKSTIIGAGLSSSEALDVLQAMETADPLOQL----- 115
 Db 518 LRKIDQE-MEGLNHTTTRTOME-MTKDKADKXDEQIKIKS--RHSDELISLQGFENK 573
 QY 116 -----VLSAVSGHLKTSDDI--LPKALASPPVIRLEAAVRLANLKNTKTYIDHLHSFIH 168
 Db 574 KQLEDMWLHSKSEKINQTRDLAKINKELAS-----SEQKNKHNNBELK 616
 QY 169 KLPEIOQLSAIF-----LRLETESDAYIRDLAAKSAISRATALQIG 214
 Db 617 RREBOLSSYEDKLPDVGCGSQDPESDLDRKKEIEKSKQ-RANLAG-ATAVYSQFTIQLT 674
 QY 215 EYQOKRFPLTLNLLTSASPOD-----QEAIIYALGKLDGQSYNNIKQLQKPDVDVTL 269
 Db 675 DENOSCCVOCORVQTEALOEIVISDLQSKRLAPDKLKSSELSK-KKERRDMDGLV 733
 QY 270 AAAQALLAGKEEDALPYIK-----KOALEERPRALYALRHPSELGIPIAL 316
 Db 734 PMROSIIDLKEKE--IPELRNKLQNVNRDIQRLKNDIEQETLLGTI--MPEESAKVCL 789
 QY 317 P-----IFLTKNSEKALNVALALLEGCDTPKLEYIT-ERLVQHYNETALASFSK 368
 Db 790 TDVTIMERQEMELKDOVERKIAQAQAKIQ-GIDLDRTVOQVQOEKQKHLDIVSSKTEL 848
 QY 369 GRTLONNKRNITVPODPOER-ERLSTTRGLEQILFELRPLEAVL----- 416
 Db 849 NRKL-----IQDDQEDLOHKSSTTNELKSEKLOISTNLQRRQDLEQVLESTREV 898
 QY 417 PCYIKLASQKTQ--LATTAISFLSHTSHQDALDLFOAKLPGEPPIIRAVADLAIVN 473
 Db 899 QSLYREIKDAKEQVSPLETTLEKF-----QOEKEBELINKK-----NTSNKIAQOKL 944
 QY 474 TKDPEKRSLLDYAKKLIQETLLFVDTENQRPHEMPRLROYVTPESRTKIMLDILETLA 533
 Db 945 NDIEKAVNHHGYMK-----DIENYIQDGKDIYKQKETELNKNVIAQLSECEK-H 993
 QY 534 TSKSESDIRLLIQ 546
 Db 994 KEKINEDMKLRQ 1006
 RESULT 11
 US-08-687-080-51
 ; Sequence 51, Application US/08687080
 ; Patent No. 5965427
 ; GENERAL INFORMATION:
 ; APPLICANT: Gregory Dolganov
 ; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
 ; NUMBER OF SEQUENCES: 175

[illegible]

	Query Match	3.9%	Score 111	DB 1	Length 2101	
	Best Local Similarity	21.9%	Pred. No. 0.85			
	Matches 113; Conservative	81;	Mismatches 184;	Indels 138;	Gaps 25;	
OY	38 OSTOQALATYIEALDAYDHDHFVLARKIGEDY-LKOSIHSSDPOTRKSTIIICAGLAGSSE	96				
Db	1387 OSKOAAAGGLRAELLRQA-----RELGGELPLRKKVAE--QERTQQQLAERKAASYE	1435				
OY	97 ALDYLSQAMETADPLOQLLVLSAVSGHLGKTSD-----DLLFRALASPYEVTRLE	146				
Db	1436 QLSMKRKHNG-----LLAEENRGSGGERANLCGRPLEVELDQAREKYVDELAAVRAD	1486				
OY	147 AAYRLANK-----NNKVIDLHLSFTIHKLPDEETOCSAAIFLRIE	186				
Db	1487 AETFLAEYQRARQSTARLEYMTAKYEGAKKYLVEERORF-----QEEROKTLTA-----Q	1536				
OY	187 TEESDAYIRDLIAAKSAIRSATALQT--GEYQO--KRFLPTLRNLITSASPDOODEAILY	242				

```

Db 1537 VEBSSKLLADSDASVVOQQKLLKVAQAGSGESQGEQAEFRQQAQNLNQLAQNLQSOKEBAAEHY 1556
QY 243 ALGKAKGOSYUAIKK-----QIQRPDVDTLAAQALIALGKEDALPYI 288
Db 1597 KL-OMEKAKHHYDAKOONOELOBLRSLDEOLKENEKELRAEAE---LGHEDQAGLK 1651
QY 289 KQALTEERPALYA-LRHLPEIG-----IPALPFLTKNSEAK--LNVALA 334
Db 1652 TKEA-EDGCHNLIAQVRSLEQAQVHAHQDLQRLQSKQVATDA-LKSREPAQKPOLDSID 1709
QY 335 LLEIGDTPKILEYTERLVQVPHYNTLALSFSKSGTLDQNMKNVNIYRPDQE--KRL 392
Db 1710 SLDSICEGTPLS-TSKLPRTOPDST-----SVGEPAASPIORL 1749
QY 393 LSTTRGIEBOILTFP--FLPKAEVLPICIKLL--ASQQTQLATVATSPLSHSHOEL 447
Db 1750 PPKVESLESLEYTPIPARSAQPLESSLDSDGVFLDSGKRTSARRRQTQIINITMKL 1809
QY 448 DLFGAAKLPGEPITIRYADALAIYNTLTKPEKKRL 483
Db 1810 DV-----EEP---DSAASTSYSTRSAPASQST 1834

```

Query Match 3.9%; Score 111; DB 1; Length 2101;
 Best Local Similarity 21.9%; Pred. No. 0.85;
 Matches 113; Conservative 81; Mismatches 184; Indels 138; Gaps 25;

```

QY 38 OSTOQALATYALDAGDHDFVLKIGEDY-LKOSIHSSDPOTPKSTITIGAGLSSE 96
DB 1387 OSKQAAAGGIRALLRQ-----RELGELLPLRKVAE---QERTAOQLRAEKASYAE 1435
QY 97 ALDVLQAMETADPLQQLLVLSAVSGHLGKTS-----DLFKALASPYVIRLE 146
DB 1436 QLSMLKKAHG-----LAEENRGGERANLGRQFLEVLDAQREKYVGEIAVRAVD 1486
QY 147 AAYRLANK-----NTKYIDHLSFIKLPBEIOLCSAIFLRLE 186
DB 1487 AETRLAEVOREASTARELEVNTAKYEGAKRVLEERORF---OBEROKLTA-----Q 1536
QY 187 TEESDAYIRDLAARKSAIRSATALQI--GEYQO--KRFLPTLRNLTSASPODEALY 242
DB 1537 VELSKKLADSOAKKVOQOKIKAVQAGGSEGOEAKRQAOQLNELQALSKOEQAHEY 1596
QY 243 ALGKLKDGOSTYNNIK-----QLOKPDVDTLAAQALIALGKEEDALPYI 288
DB 1597 KL-QMEKAKTHYDAKKOONOELQOLRSLEOLQENKELRAEAE---LGHELOQAGLK 1651
QY 289 KKOALEERPRALYA-LRHLPSRIG-----IPALPFLTKNSEAR--LNVALA 334
DB 1652 TKEA-EOTCRHLTAQVRSLEAOVAHADQOLRDGKFQVATDA-LKSREPQAPQLDLSID 1709
QY 335 LLELGGDTPKLEYITERLVOPHYNETLALSFSGRTLONMKRVNIIVPODPOE--RERL 392
DB 1710 SLDSCEGTPLS-ITSKLPTQPDGT-----SYGGEASPISQRL 1749
QY 393 LSTRGLEEQILITFL---FRLPKEAYLPCIKYL--ASOKTQLATTAISFHSHTSHEAL 447
DB 1750 PPKVESLESLEYTPIPARSOAPLESLSLSDVFLDSGRKTRRARRTQIINIMTKKL 1809
QY 448 DLFQAAKLPEPIIRAYADLAIVNLTKDPEKKRSL 483
DB 1810 DV-----EPP---DSANSSFYSTRSAPASQASL 1834

```

RESULT 15
 US-08-483-924-4
 ; Sequence 4, Application us/08483924
 ; Patent No. 5882876
 ; GENERAL INFORMATION:
 ; APPLICANT: TOUTATLY, GARY
 ; APPLICANT: LIDGARD, GRAHAM P
 ; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
 ; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
 ; STREET: 125 HIGH STREET
 ; CITY: BOSTON
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/483,924
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: PITCHER ESQ, EDMUND R
 ; REGISTRATION NUMBER: 27,829
 ; REFERENCE/DOCKET NUMBER: MTP-013

TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 248-7000
 ; TELEFAX: (617) 248-7100
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2101 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-483-924-4

Query Match 3.9%; Score 111; DB 2; Length 2101;
 Best Local Similarity 21.9%; Pred. No. 0.85;
 Matches 113; Conservative 81; Mismatches 184; Indels 138; Gaps 25;

```

QY 38 OSTOQALATYALDAGDHDFVLKIGEDY-LKOSIHSSDPOTPKSTITIGAGLSSE 96
DB 1387 OSKQAAAGGIRALLRQ-----RELGELLPLRKVAE---QERTAOQLRAEKASYAE 1435
QY 97 ALDVLQAMETADPLQQLLVLSAVSGHLGKTS-----DLFKALASPYVIRLE 146
DB 1436 QLSMLKKAHG-----LAEENRGGERANLGRQFLEVLDAQREKYVGEIAVRAVD 1486
QY 147 AAYRLANK-----NTKYIDHLSFIKLPBEIOLCSAIFLRLE 186
DB 1487 AETRLAEVOREASTARELEVNTAKYEGAKRVLEERORF---OBEROKLTA-----Q 1536
QY 187 TEESDAYIRDLAARKSAIRSATALQI--GEYQO--KRFLPTLRNLTSASPODEALY 242
DB 1537 VELSKKLADSOAKKVOQOKIKAVQAGGSEGOEAKRQAOQLNELQALSKOEQAHEY 1596
QY 243 ALGKLKDGOSTYNNIK-----QLOKPDVDTLAAQALIALGKEEDALPYI 288
DB 1597 KL-QMEKAKTHYDAKKOONOELQOLRSLEOLQENKELRAEAE---LGHELOQAGLK 1651
QY 289 KKOALEERPRALYA-LRHLPSRIG-----IPALPFLTKNSEAR--LNVALA 334
DB 1652 TKEA-EOTCRHLTAQVRSLEAOVAHADQOLRDGKFQVATDA-LKSREPQAPQLDLSID 1709
QY 335 LLELGGDTPKLEYITERLVOPHYNETLALSFSGRTLONMKRVNIIVPODPOE--RERL 392
DB 1710 SLDSCEGTPLS-ITSKLPTQPDGT-----SYGGEASPISQRL 1749
QY 393 LSTRGLEEQILITFL---FRLPKEAYLPCIKYL--ASOKTQLATTAISFHSHTSHEAL 447
DB 1750 PPKVESLESLEYTPIPARSOAPLESLSLSDVFLDSGRKTRRARRTQIINIMTKKL 1809
QY 448 DLFQAAKLPEPIIRAYADLAIVNLTKDPEKKRSL 483
DB 1810 DV-----EPP---DSANSSFYSTRSAPASQASL 1834

```

Search completed: August 20, 2002, 04:11:08
 Job time: 3293 sec